CONFORMAL WASSERSTEIN DISTANCE:
II. COMPUTATIONAL ASPECTS AND EXTENSIONS

Y. LIPMAN, J. PUENTE, I. DAUBECHIES

ABSTRACT. This paper is a companion paper to [19]. We provide numerical procedures and algorithms for computing the alignment of and distance between two disk type surfaces. We provide a convergence analysis of the discrete approximation to the arising mass-transportation problems. We furthermore generalize the framework to support sphere-type surfaces, and prove a result connecting this distance to local geodesic distortion. Lastly, we provide numerical experiments on several surfaces' datasets and compare to state of the art method.

1. INTRODUCTION AND BACKGROUND

Alignment of surfaces plays a role in a wide range of scientific disciplines. It is a standard problem in comparing different scans of manufactured objects; various algorithms have been proposed for this purpose in the computer graphics literature. It is often also a crucial step in a variety of problems in medicine and biology; in these cases the surfaces tend to be more complex, and the alignment problem may be harder. For instance, neuroscientists studying brain function through functional Magnetic Resonance Imaging (fMRI) typically observe several people performing identical tasks, obtaining readings for the corresponding activity in the brain cortex of each subject. In a first approximation, the cortex can be viewed as a highly convoluted 2-dimensional surface. Because different cortices are folded in very different ways, a synthesis of the observations from different subjects must be based on appropriate mappings between pairs of brain cortex surfaces, which reduces to a family of surface alignment problems [9, 28]. In another example, paleontologists studying molar teeth of mammals rely on detailed comparisons of the geometrical features of the tooth surfaces to distinguish species or to determine similarities or differences in diet [2].

Mathematically, the problem of surface alignment can be described as follows: given two 2-surfaces $\mathcal{M}$ and $\mathcal{N}$, find a mapping $f : \mathcal{M} \to \mathcal{N}$ that preserves, as best possible, "important properties" of the surfaces. The nature of the "important properties" depends on the problem at hand. In this paper, we concentrate on preserving the geometry, i.e., we would like the map f to preserve intrinsic distances, to the extent possible. In terms of the examples listed above, this is the criterion traditionally selected in the computer graphics literature; it also corresponds to the point of view of paleontologists studying tooth surfaces. To align cortical surfaces, one typically uses the Talairach method [18] (which relies on geometrically defined landmarks and is thus geometric in nature as well), although alignment based on functional correspondences has been proposed more recently [28].

In [19] a novel procedure between disk type surfaces was proposed, based on uniformization theory and optimal mass transportation. In a nutshell, the method maps two surfaces $\mathcal{M}, \mathcal{N}$ to densities μ, ν (interpreted as mass densities) defined on the hyperbolic disk $\mathcal{D} = \{z \mid |z| < 1\}$, their canonical uniformization space. (Apart from simplifying the description of the surface, this also removes any effect of global translations and rotations on the description of each individual surface.) The alignment problem can then be studied in the framework of Kantorovich mass-transportation [16] between these metric densities. Mass-transportation seeks to minimize the "average distance" over which mass needs to be "moved" (in the most efficient such moving procedure) to transform one mass density μ into another, ν:

$$(1.1) \qquad T_c(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathcal{D} \times \mathcal{D}} c(z, w) d\pi(z, w),$$

1

where $c(z, w) \geq 0$ is a cost function, and $\Pi(\mu, \nu)$ is the collection of probability measures on $\mathcal{D} \times \mathcal{D}$ with marginals μ and ν (resp.), that is, for $A \subset \mathcal{D}$, $B \subset \mathcal{D}$, $\pi(A \times \mathcal{D}) = \mu(A)$ and $\pi(\mathcal{D} \times B) = \nu(B)$.

In our case the uniformizing metric density (or conformal factor) corresponding to an initial surface is not unique, but is defined only up to a Möbius transformation. Because a naïve application of mass-transportation on the hyperbolic disk would not possess the requisite invariance under Möbius transformations, we generalize the mass-transportation framework, and replace the cost function $d(z, w)$ traditionally used in defining the "average displacement distance" by a cost that depends on μ and ν, $d^R_{\mu,\nu}(z, w)$, measuring the dissimilarity between the two metric densities on R-neighborhoods of z and w (where R is a parameter that controls the size of the neighborhoods):

$$(1.2) \qquad T^R_d(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathcal{D} \times \mathcal{D}} d^R_{\mu,\nu}(z, w) d\pi(z, w).$$

Introducing neighborhoods also makes the definition less sensitive to noise in practical applications. The optimal way of transporting mass in this generalized framework defines a corresponding optimal way of aligning the surfaces. This approach also allows us to define a new distance, $\mathbf{d}^R(\mathcal{M}, \mathcal{N})$, between surfaces; the average distance over which mass needs transporting (to transform one metric density into the other) quantifies the extent to which the two surfaces differ.

This paper contains three contributions that complement and extend [19]. The first of these is to provide an algorithm for approximating $\mathbf{d}^R(\mathcal{M}, \mathcal{N})$ and to prove a convergence result for this algorithm. In order to state this goal more precisely, we introduce some technicalities and notations now.

1.1. Uniformization.

Uniformization theory for Riemann surfaces [31, 14] allows conformally flattening disk type surfaces onto the unit disk $\phi : \mathcal{M} \to \mathcal{D}$ in $\mathbb{C}$, where ϕ is the conformal flattening map, $\mathcal{D} = \{z \mid |z| < 1\}$ is the unit disk, and the disk coordinate system is denoted by $z = x^1 + \mathrm{i}x^2$. The surface's Riemannian metric g is then pushed-forward to a diagonal metric tensor

$$\widetilde{g} = \phi_* g = \mu^E(z) \, \delta_{ij} \, dx^i \otimes dx^j,$$

where $\mu^E(z) > 0$, Einstein summation convention is used, and the subscript $*$ denotes the "push-forward" action; the superscript E stands for Euclidean. The function μ^E can also be viewed as the *density function* of the measure $\mathrm{vol}_\mathcal{M}$ induced by the Riemann volume element: for (measurable) $A \subset \mathcal{M}$,

$$(1.3) \qquad \mathrm{vol}_\mathcal{M}(A) = \int_{\phi(A)} \mu^E(z) \, d\mathrm{vol}_E(z),$$

where $d\mathrm{vol}_E(z) = dx^1 \wedge dx^2$ is the Euclidean area element . For a second surface $\mathcal{N}$ with Riemannian metric h we will denote its pushed-forward metric on the uniformization disk $\mathcal{D}$ by $\widetilde{h} = \phi_* h = \nu^E(w) \, \delta_{ij} \, dy^i \otimes dy^j$, where the coordinates in the unit disk are $w = y^1 + \mathrm{i}y^2$.

We use the hyperbolic metric on the unit disk $(1 - |z|^2)^{-2} \delta_{ij} dx^i \otimes dx^j$ as a reference metric; the surface density w.r.t. the hyperbolic metric (conformal scaling) is

$$(1.4) \qquad \mu^H(z) := (1 - |z|^2)^2 \, \mu^E(z),$$

where the superscript H stands for hyperbolic.

We shall often drop this superscript: unless otherwise stated $\mu = \mu^H$, and $\nu = \nu^H$ in what follows. The density function $\mu = \mu^H$ satisfies

$$\mathrm{vol}_\mathcal{M}(A) = \int_{\phi(A)} \mu(z) \, d\mathrm{vol}_H(z),$$

where $d\mathrm{vol}_H(z) = (1 - |z|^2)^{-2} \, d\mathrm{vol}_E(z)$. We will use the notations μ, ν also to represent the *measures* $\mathrm{vol}_\mathcal{M}, \mathrm{vol}_\mathcal{N}$ (resp.), where the exact meaning will be clear from the context.

The conformal mappings of $\mathcal{D}$ to itself are the disk-preserving Möbius transformations, they constitute the group $M_\mathcal{D}$ of isometries of the hyperbolic disk. An arbitrary element $m \in M_\mathcal{D}$ is characterized by three real

parameters:

$$m(z) = e^{\mathbf{i}\theta}\frac{z-a}{1-\bar{a}z}, \ a \in \mathcal{D}, \ \theta \in [0, 2\pi). \tag{1.5}$$

The pull-back $m^*\mu(z)$ and the push-forward $m_*\mu(w)$ of the density μ by the Möbius transformation m are given by the formulas

$$m^*\mu(z) = \mu(m(z)), \tag{1.6}$$

and

$$m_*\mu(w) = \mu(m^{-1}(w)), \tag{1.7}$$

respectively.

It follows that checking whether or not two surfaces $\mathcal{M}$ and $\mathcal{N}$ are isometric, or searching for isometries between $\mathcal{M}$ and $\mathcal{N}$, is greatly simplified by considering the conformal mappings from $\mathcal{M}$, $\mathcal{N}$ to $\mathcal{D}$: once the (hyperbolic) density functions μ and ν are known, it suffices to identify $m \in M_\mathcal{D}$ such that $\nu(m(z))$ equals $\mu(z)$ (or "nearly" equals, in a sense to be made precise).

1.2. **Optimal volume transportation for surfaces.** To adapt the optimal transportation framework to the alignment of surfaces, we use an isometry invariant cost function $d_{\mu,\nu}^R(z, w)$ that is plugged into the transportation framework (1.1). This special cost function $d_{\mu,\nu}^R(z, w)$ indicates the extent to which a neighborhood of the point z in $(\mathcal{D}, \mu)$, the (conformal representation of the) first surface, is isometric with a neighborhood of the point w in $(\mathcal{D}, \nu)$, the (conformal representation of the) second surface. Two definitions are in order: 1) the neighborhoods we will use, and 2) how we characterize the (dis)similarity of two neighborhoods, equipped with different metrics.

For the neighborhoods, we take the hyperbolic geodesic disks $\Omega_{z_0,R}$ of radius R, where we let z_0 range over $\mathcal{D}$, but keep the radius $R > 0$ fixed. The following gives an easy procedure to construct these disks. If $z_0 = 0$, then the hyperbolic geodesic disks centered at $z_0 = 0$ are also "standard" (i.e. Euclidean) disks centered at 0: $\Omega_{0,R} = \{z \, ; \, |z| \leq r_R\}$, where $r_R = \tanh(R)$. The hyperbolic disks around other centers are images of these central disks under Möbius transformations (= hyperbolic isometries): setting $m(z) = (z - z_0)(1 - z\bar{z}_0)^{-1}$, we have

$$\Omega_{z_0,R} = m^{-1}(\Omega_{0,R}). \tag{1.8}$$

Next, the (dis)similarity of the pairs $(\Omega_{z_0,R}, \mu)$ and $(\Omega_{w_0,R}, \nu)$ is defined via pull-back of ν and using the standard induced norm (see [19] for more details). The final cost function is achieved by taking the infimum over all Möbius transformations m such that $m(z) = w$:

$$d_{\mu,\nu}^R(z_0, w_0) := \inf_{m \in M_\mathcal{D} \, , \, m(z_0)=w_0} \int_{\Omega_{z_0,R}} |\, \mu(z) - (m^*\nu)(z) \,| \, d\mathrm{vol}_H(z), \tag{1.9}$$

where $d\mathrm{vol}_H(z) = (1 - |z|^2)^{-2} \, dx^1 \wedge dx^2$ is the volume form for the hyperbolic disk.

As proved in [19] $d_{\mu,\nu}^R(\cdot, \cdot)$ is a metric on $\mathcal{D}$ and as a consequence

$$\mathbf{d}^R(\mathcal{M}, \mathcal{N}) = T_d^R(\mu, \nu) \tag{1.10}$$

(with $T_d^R(\mu, \nu)$ as defined in 1.2) defines a semi-metric in the space of disk-type surfaces. To ensure that this is a metric rather than only a semi-metric, we add an extra assumption, namely that no (orientation-preserving) self-isometries exist within each of the compared surfaces. For discussion and more detail we refer the reader to [19].

1.3. **Overview.** We can now formulate a precise overview of the algorithm for approximating $\mathbf{d}^R(\mathcal{M}, \mathcal{N})$, and discuss its convergence properties.

In a nutshell, the key steps of the algorithm are: 1) approximate uniformization for piecewise linear surface representations, 2) discretize the continuous measures μ, ν based on discrete sample sets $Z = \{z_i\}_{i=1}^n$, $W = \{w_j\}_{j=1}^p$, obtaining discrete measures by μ_Z, ν_W (resp.), 3) approximate the measure-dependent cost function $\breve{d}_{\mu,\nu}^R(z, w) \approx d_{\mu,\nu}^R(z, w)$, and 4) calculate the discrete optimal transportation cost $T_{\breve{d}}^R(\mu_Z, \nu_W)$ between the discrete measures μ_Z, ν_W based on the approximated cost function $\breve{d}_{\mu,\nu}^R(z, w)$.

In the heart of our analysis we prove the convergence $T_{\breve{d}}^R(\mu_Z, \nu_W) \to \mathbf{d}^R(\mathcal{M}, \mathcal{N})$ as the "mesh size" of the samplings used in steps 2, 3 and 4, tend to zero. More precisely, we define the *fill distance* $\varphi_g(Z)$ for the metric tensor g and the sample set Z as

$$(1.11) \qquad \varphi_g(Z) := \sup \left\{ r > 0 \mid z \in \mathcal{M} : B_g(z, r) \cap Z = \emptyset \right\},$$

where $B_g(z, r)$ is the geodesic open ball of radius r centered at z. That is, $\varphi_g(Z)$ is the radius of the largest geodesic ball that can be fitted on the surface $\mathcal{M}$ without including any point of Z. The smaller $\varphi_g(Z)$, the finer the sampling set. We prove the following theorem:

Theorem 1.1. *Let μ, ν be Lipschitz continuous probability densities (w.r.t the hyperbolic measure) defined over $\mathcal{D}$. Let $\mathcal{M}, \mathcal{N}$ be the disk-type surfaces defined by the metric tensors $g = \mu(z)(1 - |z|^2)^{-2}\delta_{ij}dx^i \otimes dx^j$, $h = \nu(w)(1 - |w|^2)^{-2}\delta_{ij}dy^i \otimes dy^j$ (resp.), let Z, W be discrete sample sets on $\mathcal{M}, \mathcal{N}$ (resp.), $\{p_k\}$ samples in $\mathcal{D}$ used for numerical integration, and let L be the number of uniformly spaced points on the circle used to define $T_{\breve{d}}^R$ (see below). Then*

$$\begin{aligned} \left| \mathbf{d}^R(\mathcal{M}, \mathcal{N}) - T_{\breve{d}}^R(\mu_Z, \nu_W) \right| &= \left| T_{\breve{d}}^R(\mu, \nu) - T_{\breve{d}}^R(\mu_Z, \nu_W) \right| \\ &\leq \omega_{d_{\mu,\nu}^R}\left(2\max(\varphi_g(Z), \varphi_h(W))\right) + C_1 \varphi_E(\{p_k\}) + C_2 L^{-1} + \epsilon_{\text{unif}}, \end{aligned}$$

where

$$\omega_{d_{\mu,\nu}^R}(t) = \sup_{d_{\mathcal{M}}(z,z') + d_{\mathcal{N}}(w,w') < t} \left| d_{\mu,\nu}^R(z, w) - d_{\mu,\nu}^R(z', w') \right|.$$

denotes the modulus of continuity *of the cost function $d_{\mu,\nu}^R$, and C_1, C_2 are constants that depend only upon μ, ν, R.*

Here $\varphi_E(\{p_k\})$ and L^{-1} are two algorithm parameters that can be made arbitrary small. The error term ϵ_{unif} concerns only the approximations made in the discrete uniformizations for each of the two surfaces, separately; we'll come back to it below – suffice it to say here that it is much smaller than the other error terms, in practice. Finally, it was proved in [19] that the cost function $d_{\mu,\nu}^R(z, w)$ is uniformly continuous on $\overline{\mathcal{D}} \times \overline{\mathcal{D}}$ and therefore $\omega_{d_{\mu,\nu}^R}\left(2\max(\varphi_g(Z), \varphi_h(W))\right) \to 0$ as the fill-distances of the sets Z, W go to zero.

Two comments are in order: first, we believe that the cost function $d_{\mu,\nu}^R(z, w)$ is Lipschitz rather than just uniformly continuous. If this is the case, then our analysis guarantees linear convergence of the algorithm (leaving aside the ϵ_{unif} term). We leave checking the precise regularity of the cost function to future work. Second, we should discuss in more detail ϵ_{unif}, containing the errors produced from the discrete uniformization. One typically starts from a piecewise flat approximation of the surfaces $\mathcal{M}, \mathcal{N}$, given by triangle meshes with a very fine mesh size, providing much finer sampling than the Z or W used in steps 2 through 4. (For instance, in our numerical computations, the parameter L and the sample sets Z, W were picked so that $L^{-1}, \varphi_g(Z), \varphi_h(W)$ had magnitude .02,.06,.06 (resp.); the mesh size in the original triangulation of $\mathcal{M}, \mathcal{N}$ is of order .01.) How to construct discrete approximations of the uniformization of the surfaces, starting from these approximations to $\mathcal{M}, \mathcal{N}$, is a research area in its own right, and several different methods have been proposed [12, 26, 30]; in our work we adopt the approach of [26, 20]. The error ϵ_{unif} contributed by this component of our algorithm is governed by the difference between the "true" μ, ν and the $\mu^{approx}, \nu^{approx}$ stemming from the discrete triangulation approximation, followed by the discrete uniformization, and is

bounded by (using the triangle inequality proved in Theorem 3.11 in [19])

$$\epsilon_{\text{unif}} \leq T_d^R(\mu, \mu^{approx}) + T_d^R(\nu, \nu^{approx}).$$

We expect this difference to be proportional to the triangulation mesh size, and to be negligible compared to the errors we analyze explicitly in this paper. Since the convergence analysis of discrete uniformization has not settled yet into its final form, and given the much smaller size of this component of the error (both expected and borne out by numerical experiments), we have opted here to view the discrete uniformization as a "black box", the analysis of which is outside the scope of this paper, and to neglect this part of the error. We concern ourselves here with the error made by our algorithm in the approximation to $T_d^R(\mu^{approx}, \nu^{approx})$, namely with

$$\left| T_d^R(\mu^{approx}, \nu^{approx}) - T_d^R(\mu_Z, \nu_W) \right|.$$

We now turn to the other contributions made by this paper. In an earlier version of the paper, Theorem 2.4 was the main result. Interesting questions and challenges by the reviewers led us to investigate extensions and further mathematical properties of our construction; the results are formulated as two further contributions.

The first of these is a generalization of the framework above to other types of surfaces. We show how a similar distance $\mathbf{d}^A(\mathcal{M}, \mathcal{N})$ can be defined for genus-zero, or sphere-type surfaces. This involves some new ideas, since the absence of a distance function on the sphere S^2 that would be invariant under all conformal maps from S^2 to itself, implies that the definitions of the neighborhoods $\Omega_{z_0, R}$ cannot simply be copied from the case for disk-type surfaces.

The final contribution of this paper concerns possible connections between the distance $\mathbf{d}^R(\mathcal{M}, \mathcal{N})$ and the notion of geodesic distortion. Although there is certainly much more to be said upon this topic than we do here, we do present a first result, showing that if the distance $\mathbf{d}^R(\mathcal{M}, \mathcal{N})$ between two disk-like surfaces is small, then the two surfaces are *locally* near-isometric. More precisely, we prove the following:

Theorem 1.2. *Let $\mathcal{M}$ and $\mathcal{N}$ be differentiable disk-like surfaces. If $\mathbf{d}^R(\mathcal{M}, \mathcal{N})$ is sufficiently small, then we can cover $\mathcal{M}$ (minus an arbitrarily small boundary layer) with patches $\Omega_{z_0, R}$ and define mappings $f^{z_0} : \mathcal{M} \to \mathcal{N}$ (Möbius transformations) such that for all $z_1, z_2 \in \mathcal{M}$ (not very close to the boundary) with geodesic distance $d_g(z_1, z_2) \leq r(R)$, $r(R) > 0$, there exists a patch $\Omega_{z_0, R}$ such that $z_1, z_2 \in \Omega_{z_0, R}$, and*

$$(1.12) \qquad \left(1 - C_2 \mathbf{d}^R(\mathcal{M}, \mathcal{N})^{1/3} \right) d_g(z_1, z_2) \leq d_h(f^{z_0}(z_1), f^{z_0}(z_2)) \leq \left(1 + C_1 \mathbf{d}^R(\mathcal{M}, \mathcal{N})^{1/3} \right) d_g(z_1, z_2),$$

where $d_g(z_1, z_2), d_h(w_1, w_2)$ are the geodesic distances of $z_1, z_2 \in \mathcal{M}$ and $w_1, w_2 \in \mathcal{N}$, respectively, and $C_1, C_2 > 0$ are constants independent of the choice of z_1, z_2.

1.4. **Related work.** The approach taken in this paper is related to the computer graphics constructions in [20], which rely on the representation of isometries between topologically equivalent simply-connected surfaces by Möbius transformations between their uniformization spaces, and which exploit that 1) the Möbius group has small dimensionality (e.g. 3 for disk-type surfaces and 6 for sphere-type) and 2) changing the metric in one piece of a surface has little influence on the uniformization of distant parts. These two observations lead, in [20], to fast and effective algorithms to identify near-isometries between differently deformed versions of a surface. In our present context, these same observations lead to a simple algorithm for surface alignment, reducing it to a linear programming problem.

Other distances between surfaces have been used recently for several applications [21]. A prominent mathematical approach to define distances between surfaces considers the surfaces as special cases of *metric spaces*, and uses then the Gromov-Hausdorff (GH) distance between metric spaces [11]. The GH distance between metric spaces X and Y is defined through examining all the isometric embedding of X and Y into (other) metric spaces; although this distance possesses many attractive mathematical properties, it is inherently hard computationally [22, 1]. For instance, computing the GH distance is equivalent to a non-convex quadratic programming problem; solving this directly for correspondences is equivalent to integer quadratic assignment, and is thus NP-hard [6]. In addition, the non-convexity implies that the solution found in practice may be

a local instead of a global minimum, and is therefore not guaranteed to give the correct answer for the GH distance. The distance between surfaces as we define in [19] does not have these shortcomings because the computation of the distance between surfaces using this approach can be recast as a linear program, and can therefore be implemented using efficient polynomial algorithms that are moreover guaranteed to converge to the correct solution.

In [21], the GH distance of [22] is generalized by introducing a quadratic mass transportation scheme to be applied to metric spaces equipped with a measure (mm spaces); the computation of this Gromov-Wasserstein distance for mm spaces is somewhat easier and more stable to implement than the original GH distance [21]. A crucial aspect in which our work differs from [21] is that, in contrast to the (continuous) quadratic programming method proposed in [21] to compute the Gromov-Wasserstein distance between mm spaces, our conformal approach leads to a convex (even linear) problem, solvable via a linear programming method.

It is worth mentioning that optimal mass transportation has been used in the engineering literature as well, to define interesting metrics between images; in this context the metric is often called the Wasserstein distance. The seminal work for this image analysis approach is the paper by Rubner et al. [27], in which images are viewed as discrete measures, and the distance is called appropriately the "Earth Mover's Distance".

Another related method is presented in the papers of Zeng et al. [33, 34], which also use the uniformization space to match surfaces. Our work differs from that of Zeng et al. in that they use prescribed feature points (defined either by the user or by extra texture information) to calculate an interpolating harmonic map between the uniformization spaces, and then define the final correspondence as a composition of the uniformization maps and this harmonic interpolant. This procedure is highly dependent on the prescribed feature points, provided as extra data or obtained from non-geometric information. In contrast, our work does not use any prescribed feature points or external data, and makes use of only the geometry of the surface; in particular we utilize the conformal structure itself to define deviation from (local) isometry.

1.5. **Organization.** Section 2 presents the main steps for the discretization of the continuous case and provides algorithmic aspects for the alignment procedure. Section 3 generalizes the method to sphere-type surfaces. Section 4 provides a theoretical result connecting our distance directly to local geodesic distortion. Section 5 presents experimental validation of the algorithms and concludes; in particular, we report results of the method applied to various benchmark data sets and provide a comparison to a state-of-the-art method. This paper also contains four appendices: A) contains few approximation results used by our algorithm, B) contains background on the discrete conformal mapping we use, C) contains proofs of some properties of the linear program solution, and D) contains the approximation analysis of the discrete optimal transport cost to its continuous counterpart.

2. ALGORITHM FOR COMPARING DISK-TYPE SURFACES AND ANALYSIS

Transforming the theoretical framework discussed above into an algorithm requires several steps of approximation. Our general plan is to recast the transportation eq. (1.2) as a linear programming problem between discrete measures. The steps of our algorithm are as follows:
Preprocess: approximating the surfaces' uniformization,
Step 1: discretizing the resulting continuous measures,
Step 2: approximating the cost function $d_{\mu,\nu}^R(\cdot, \cdot)$,
Step 3: solving a linear programming problem to achieve the final approximation of the distance, and the optimal transportation plan (correspondences).
Step 4 (optional): extract a consistent set of correspondences.

In the following we describe in detail each of these steps; we also provide a convergence analysis for steps 1-3, but not for the *preprocess step*. (As explained in the introduction the convergence of the approximated

uniformization is not the focus of this paper, and we consider it as a "black box".) For the sake of completeness, and as a guide to readers who would like to implement the algorithm, we nevertheless include a description of this part in Appendix B.

2.1. **Step 1: Discretizing continuous measures.** In this subsection we indicate how to construct the discrete measures μ_Z, ν_W used in further steps.

Given the measure $\mu = \mathrm{vol}_{\mathcal{M}}$ on $\mathcal{D}$, we discretize it by first distributing n points $Z = \{z_i\}_{i=1}^n$ "uniformly w.r.t. μ". Details on the particular algorithm we used for sampling are provided in Appendix B (we use the same technique for the sets Z, W described there). For $i = 1, \ldots, n$, we define the sets $\{\Xi_i\}_{i=1}^n$ to be the Voronoi cells corresponding to $z_i \in Z$ defined by the metric of $\mathcal{M}$; this gives a partition of $\mathcal{D}$ into disjoint sets, $\mathcal{D} = \cup_{i=1}^n \Xi_i$. For a more detailed definition of Voronoi cells as-well as properties of the discrete measures see Appendix D. Next, define the discrete measure μ_Z as a superposition of delta measures localized in the points of Z, with weights given by the areas of Ξ_i, i.e.

$$(2.1) \qquad \mu_Z = \sum_{i=1}^n \mu_i \delta_{z_i},$$

with $\mu_i := \mathrm{vol}_{\mathcal{M}}(\Xi_i) = \int_{\Xi_i} \mu(z) d\mathrm{vol}_H$. Similarly we denote by $W = \{w_j\}_{j=1}^p$, $\{\Upsilon_j\}_{j=1}^p$, ν_W, and $\nu_j := \mathrm{vol}_{\mathcal{N}}(\Upsilon_j)$ the corresponding quantities for the measure $\nu = \mathrm{vol}_{\mathcal{N}}$.

We shall always assume that the surfaces $\mathcal{M}$ and $\mathcal{N}$ have the same area, which, for convenience, we can take to be 1. It then follows that the discrete measures μ_Z and ν_W have equal total mass (regardless of whether $n = p$ or not). The approximation algorithm will compute optimal transport for the discrete measures μ_Z and ν_W; the corresponding discrete approximation to the distance between $\mathcal{M}$ and $\mathcal{N}$ is then given by $T_d^R(\mu_Z, \nu_W)$.

The convergence analysis we present will be in terms of the *fill distance* $\varphi_g(Z), \varphi_h(W)$ defined in the introduction. Note that our analysis will work with any point sample sets as long as their fill distances converge to zero.

2.2. **Step 2: approximating the cost function $d_{\mu,\nu}^R$.** In order to approximate $T_d^R(\mu_Z, \nu_W)$ we need to approximate the cost function $d_{\mu,\nu}^R(z, w)$ between pairs of points $(z_i, w_j) \in Z \times W$.

Applying (1.9) to the points z_i, w_j we have:

$$(2.2) \qquad d_{\mu,\nu}^R(z_i, w_j) = \min_{m(z_i)=w_j} \int_{\Omega_{z_i,R}} \Big| \mu(z) - \nu(m(z)) \Big| d\mathrm{vol}_H.$$

To obtain $d_{\mu,\nu}^R(z_i, w_j)$ we will thus need to approximate integrals over hyperbolic disks of radius R, which is done via a separate approximation procedure, set up once and for all in a preprocessing step at the start of the algorithm.

By using a Möbius transformation $\widetilde{m}$ such that $\widetilde{m}(0) = z_0$, and the identity

$$\int_{\Omega_{z_0,R}} \Big| \mu(z) - \nu(m(z)) \Big| d\mathrm{vol}_H(z) = \int_{\Omega_{0,R}} \Big| \mu(\widetilde{m}(u)) - \nu(m \circ \widetilde{m}(u)) \Big| d\mathrm{vol}_H(u) \,,$$

we can reduce the integrals over the hyperbolic disks $\Omega_{z_i,R}$ to integrals over a hyperbolic disk $\Omega_{0,R}$ centered around zero.

To approximate the integral of a continuous function f over $\Omega_{0,R}$ we then use a rectangle-type quadrature

$$\int_{\Omega_0} f(z) \, d\mathrm{vol}_H(z) \approx \sum_k \alpha_k f(p_k)$$

where $p_k \in \Omega_{0,R}, \alpha_k \in \mathbb{R}, k = 1..K$, are the centers and coefficients (resp.) of the quadrature. The coefficients α_k are defined as the hyperbolic area of the Euclidean Voronoi cells Δ_k corresponding to the centers p_k.

We thus have the following approximation:

$$d_{\mu,\nu}^R(z_i, w_j) = \min_{m(z_i)=w_j} \int_{\Omega_{z_i,R}} \left| \mu(z) - \nu(m(z)) \right| d\text{vol}_H(z)$$

$$= \min_{m(z_i)=w_j} \int_{\Omega_{0,R}} \left| \mu(\widetilde{m}_i(z)) - \nu(m(\widetilde{m}_i(z))) \right| d\text{vol}_H(z)$$

$$(2.3) \qquad \approx \min_{m(z_i)=w_j} \sum_k \alpha_k \left| \mu(\widetilde{m}_i(p_k)) - \nu(m(\widetilde{m}_i(p_k))) \right| ,$$

where the Möbius transformations $\widetilde{m}_i$, mapping 0 to z_i, are selected as soon as the z_i themselves have been picked, and remain the same throughout the remainder of the algorithm.

Let us denote this approximation by

$$\widehat{d}_{\mu,\nu}^R(z_i, w_j) = \min_{m(z_i)=w_j} \sum_k \alpha_k \left| \mu(\widetilde{m}_i(p_k)) - \nu(m(\widetilde{m}_i(p_k))) \right|.$$

It can be shown that picking a set of centers $\{p_k\}$ with Euclidean fill-distance $\varphi_E(\{p_k\}) = h > 0$ (that is, we use the Euclidean metric to define the fill distance of the set $\{p_k\}$) leads to an $O(h)$ approximation; in Appendix A we prove:

Theorem 2.1. *For Lipschitz continuous μ, ν,*

$$\left| d_{\mu,\nu}^R(z_i, w_j) - \widehat{d}_{\mu,\nu}^R(z_i, w_j) \right| \leq C\, \varphi_E(\{p_k\}) \ ,$$

where the constant C depends only on μ, ν, R.

In practice, the minimization over $M_{\mathcal{D},z_i,w_j}$ (the set of all Möbius transformations that map z_i to w_k) in the computation of $\widehat{d}_{\mu,\nu}^R$ is discretized as well: instead of minimizing over all $M_{\mathcal{D},z_i,w_j}$, we minimize over only the Möbius transformations $\left(m_{z_i,w_j,2\pi\ell/L}\right)_{\ell=0,1,..,L-1}$, defined by

$$(2.4) \qquad\qquad\qquad m_{z_i,w_j,2\pi\ell/L} = \widetilde{m}_j \circ \mathfrak{R}_\ell \circ \widetilde{m}_i^{-1},$$

with $\widetilde{m}_i$ as defined above, $\mathfrak{R}_\ell(z) = e^{\mathbf{i}2\pi\ell/L} z$, L a parameter that reflects how many points we use to discretize $[0, 2\pi)$, and $\widetilde{m}_j \in M_{\mathcal{D}}$ an arbitrary but fixed Möbius map that takes 0 to w_j.

Taking this into account as well, we have thus

$$(2.5) \qquad d_{\mu,\nu}^R(z_i, w_j) \approx \breve{d}_{\mu,\nu}^R(z_i, w_j) := \min_{\ell=1..L} \sum_k \alpha_k \left| \mu(\widetilde{m}_i(p_k)) - \nu(m_{z_i,w_j,2\pi\ell/L}(\widetilde{m}_i(p_k))) \right| ;$$

as we prove in Appendix A the error made in approximation (2.5) is

Theorem 2.2. *For Lipschitz continuous μ, ν,*

$$\left| d_{\mu,\nu}^R(z_i, w_j) - \breve{d}_{\mu,\nu}^R(z_i, w_j) \right| \leq C_1\, \varphi_E(\{p_k\}) + C_2 L^{-1} \ ,$$

where the constants C_1, C_2 depends only on μ, ν, R.

2.3. *Step 3*: **Solving a linear program.** We now have in place all the ingredients to formulate the final linear programming problem, the solution of which approximates the distance $\mathbf{d}^R(\mathcal{M}, \mathcal{N})$. The final step is to solve a discrete optimal transportation problem between the discrete measures μ_Z and ν_W with the approximated cost function $\breve{d}_{\mu,\nu}^R(z_i, w_j)$:

$$(2.6) \qquad\qquad\qquad \sum_{i,j} \breve{d}_{ij}^R \pi_{ij} \rightarrow \min$$

$$(2.7) \qquad\qquad\qquad \begin{cases} \sum_i \pi_{ij} = \nu_j \\ \sum_j \pi_{ij} = \mu_i \\ \pi_{ij} \geq 0 \end{cases} ,$$

where $\mu_i = \mu(\Xi_i)$ and $\nu_j = \nu(\Upsilon_j)$, and $\breve{d}^R_{ij} = \breve{d}^R_{\mu,\nu}(z_i, w_j)$.

The optimal transportation plan π^* then furnishes our final approximation: $T_{\breve{d}}(\mu_Z, \nu_W) = \sum_{ij} \breve{d}^R_{ij} \pi^*_{ij}$. The approximation result will be expressed in terms of the modulus of continuity of our cost function: $\omega_{d^R_{\mu,\nu}}$. Our result will use the following regularity theorem of mass transportation, proved in Appendix D,

Theorem 2.3. *Suppose $c : \mathcal{X} \times \mathcal{Y} \longrightarrow \mathbb{R}_+$ is a continuous function, with $\mathcal{X}, \mathcal{Y}$ compact complete separable metric spaces, S and T are sample sets in $\mathcal{X}, \mathcal{Y}$ (resp.), μ, ν are probability measures on $\mathcal{X}, \mathcal{Y}$.*
(A) if c is uniformly continuous then

$$T_c(\mu_S, \nu_T) \to T_c(\mu, \nu), \quad as \; h \to 0,$$

(B) if c is Lipschitz continuous with a constant λ, then

$$|T_c(\mu, \nu) - T_c(\mu_S, \nu_T)| < 2\lambda h,$$

where, $h = \max\{\varphi_{\mathcal{X}}(S), \varphi_{\mathcal{Y}}(T)\}$, and μ_S, ν_T are as defined similarly to (2.1). (See Appendix D for precise definition.)

Our main approximation result is as follows:

Theorem 2.4. *Let μ, ν be Lipschitz continuous probability densities (w.r.t. the hyperbolic measure) defined over $\mathcal{D}$. Let $\mathcal{M}, \mathcal{N}$ be the disk-type surfaces defined by the metric tensors $g = \mu(z)(1-|z|^2)^{-2}\delta_{ij}dx^i \otimes dx^j$, $h = \nu(w)(1-|w|^2)^{-2}\delta_{ij}dy^i \otimes dy^j$ (resp.). Let π^* be the minimizer of the linear program defined by (2.6)-(2.7), then*

$$\left| \mathbf{d}^R(\mathcal{M}, \mathcal{N}) - T_{\breve{d}}(\mu_Z, \nu_W) \right| \leq \omega_{d^R_{\mu,\nu}} \left(2\max\left(\varphi_g(Z), \varphi_h(W)\right) \right) + C_1 \varphi_E(\{p_k\}) + C_2 \, L^{-1},$$

where $\omega_{d^R_{\mu,\nu}}$ denotes the modulus of continuity of the function $d^R_{\mu,\nu}$, C_1, C_2 are constants dependent only upon μ, ν, R.

Proof. First,

$$\left| \mathbf{d}^R(\mathcal{M}, \mathcal{N}) - T_{\breve{d}}(\mu_Z, \nu_W) \right| \leq \left| \mathbf{d}^R(\mathcal{M}, \mathcal{N}) - T_d(\mu_Z, \nu_W) \right| + \left| T_d(\mu_Z, \nu_W) - T_{\breve{d}}(\mu_Z, \nu_W) \right|$$
$$= I + II.$$

where $T_d(\mu_Z, \nu_W)$ is the optimal transport cost between the discrete measures μ_z, ν_W using the exact cost function $d^R_{\mu,\nu}(z_i, w_j)$.

In Appendix D (Theorem D.4) we prove that

$$I \leq \omega_{d^R_{\mu,\nu}}(2\max\{\varphi_g(Z), \varphi_h(W)\});$$

the result is proved in the more general context of compact complete separable metric spaces; we believe this result may be useful, independently of the remainder of this paper, to approximate optimal transport cost in more general contexts (see Appendix D for more details).

To bound II, denote by π'_{ij} the optimal plan in $T_{\breve{d}}(\mu_Z, \nu_W)$, then,

$$T_d(\mu_Z, \nu_W) - T_{\breve{d}}(\mu_Z, \nu_W) = \inf_{\pi} \sum_{i,j} \pi_{i,j} d^R_{i,j} - \sum_{i,j} \pi'_{i,j} \breve{d}^R_{i,j}$$
$$\leq \sum_{i,j} \pi'_{i,j} \left(d^R_{i,j} - \breve{d}^R_{i,j} \right)$$
$$\leq C_1 \, \varphi_E(\{p_k\}) + C_2 \, L^{-1},$$

where in the last inequality we used Theorem 2.2. The symmetric inequality can be achieved similarly. This completes the proof. $\square$

It is proved in [19] that $d_{\mu,\nu}^R(\cdot,\cdot)$ is uniformly continuous on $\overline{\mathcal{D}} \times \overline{\mathcal{D}}$. Therefore, the above theorem implies convergence of our discrete approximation. More specifically, our approximation will converge like the modulus of continuity of $d_{\mu,\nu}^R(\cdot,\cdot)$; remember that for uniformly continuous functions f, the modulus of continuity satisfies $\lim_{r\to 0} \omega_f(r) = 0$. As mentioned in the introduction, we believe that $d_{\mu,\nu}^R(\cdot,\cdot)$ is actually Lipschitz continuous, in that case the above theorem actually implies linear convergence rate. We leave the question of higher regularity of the cost function $d_{\mu,\nu}^R$ to future work.

In the remaining part of this subsection we discuss some variations and properties of the linear program formulation eq.(2.6)-(2.7). In practice, surfaces are often only partially isometric. Furthermore, the sampled points may also fail to have a good one-to-one and onto correspondence (i.e. there typically are some points in both Z and W that do not correspond well to any point in the other set). In these cases it is desirable to allow the algorithm to consider transportation plans π with marginals *smaller or equal* to μ and ν. Intuitively this means that we allow that only some fraction of the mass is transported and that the remainder can be "thrown away". This leads to the following formulation:

$$(2.8) \qquad \sum_{i,j} d_{ij}\pi_{ij} \to \min$$

$$(2.9) \qquad \begin{cases} \sum_i \pi_{ij} \le \nu_j \\ \sum_j \pi_{ij} \le \mu_i \\ \sum_{i,j} \pi_{ij} = \mathcal{Q} \\ \pi_{ij} \ge 0 \end{cases}$$

where $0 < \mathcal{Q} \le 1$ is a parameter set by the user that indicates how much mass *must* be transported, in total.

Since these equations and constraints are all linear, we have the following theorem:

Theorem 2.5. *The equations* (2.6)-(2.7) *and* (2.8)-(2.9) *admit a global minimizer that can be computed in polynomial time, using standard linear-programming techniques.*

When correspondences between surfaces are sought, i.e. when one surface is viewed as being transformed into the other, one is interested in restricting π to the class of permutation matrices instead of allowing all bistochastic matrices. (This means that each entry π_{ij} is either 0 or 1.) In this case the number of centers z_i must equal that of w_j, i.e. $n = N = p$, and it is best to pick the centers so that $\mu_i = \frac{1}{N} = \nu_j$, for all i, j. It turns out that these restrictions are sufficient to *guarantee* (without restricting the choice of π in any way) that the minimizing π is a permutation:

Theorem 2.6. *If $n = N = p$ and $\mu_i = \frac{1}{N} = \nu_j$, then*

(1) *There exists a global minimizer of* (2.6) *that is a permutation matrix.*
(2) *If furthermore $\mathcal{Q} = \frac{M}{N}$, where $M < N$ is an integer, then there exists a global minimizer of* (2.8) π *such that $\pi_{ij} \in \{0,1\}$ for each i, j.*

Remark 2.7. In the second case, where $\pi_{ij} \in \{0,1\}$ for each i, j, and $\sum_{i,j=1}^N \pi_{ij} = M$, π can still be viewed as a permutation of M objects, "filled up with zeros". That is, if the zero rows and columns of π (which must exist, by the pigeon hole principle) are removed, then the remaining $M \times M$ matrix is a permutation.

Proof. We first note that in both cases, we can simply renormalize each μ_i and ν_j by N, leading to the rescaled systems

$$(2.10) \qquad \begin{cases} \sum_i \pi_{ij} = 1 \\ \sum_j \pi_{ij} = 1 \\ \pi_{ij} \ge 0 \end{cases} \qquad\qquad \begin{cases} \sum_i \pi_{ij} \le 1 \\ \sum_j \pi_{ij} \le 1 \\ \sum_{i,j} \pi_{ij} = M \\ \pi_{ij} \ge 0 \end{cases}$$

To prove the first part, we note that the left system in (2.10) defines a convex polytope in the vector space of matrices that is exactly the Birkhoff polytope of bistochastic matrices. By the Birkhoff-Von Neumann

Theorem [17] every bistochastic matrix is a convex combination of the permutation matrices, i.e. each π satisfying the left system in (2.10) must be of the form $\sum_k c_k \tau^k$, where the τ^k are the $N!$ permutation matrices for N objects, and $\sum_k c_k = 1$, with $c_k \geq 0$. The minimizing π in this polytope for the linear functional (2.6) must thus be of this form as well. It follows that at least one τ^k must also minimize (2.6), since otherwise we would obtain the contradiction

$$(2.11) \qquad \sum_{ij} d_{ij}\pi_{ij} = \sum_k c_k \Big(\sum_{ij} d_{ij}\tau_{ij}^k \Big) \geq \min_k \Big\{ \sum_{ij} d_{ij}\tau_{ij}^k \Big\} > \sum_{i,j} d_{ij}\,\pi_{ij} \ .$$

The second part can be proved along similar steps: the right system in (2.10) defines a convex polytope in the vector space of matrices; it follows that every matrix that satisfies the system of constraints is a convex combination of the extremal points of this polytope. It suffices to prove that these extreme points are exactly those matrices that satisfy the constraints and have entries that are either 0 or 1 (this is the analog of the Birkhoff-von Neumann theorem for this case; we prove this generalization in a lemma in Appendix C); the same argument as above then shows that there must be at least one extremal point where the linear functional (2.6) attains its minimum. □

When we seek correspondences between two surfaces, there is thus no need to *impose* the (very nonlinear) constraint on π that it be a permutation matrix; one can simply use a (standard) linear program and Theorem 2.6 then guarantees that the minimizer for the "relaxed" problem (2.6)-(2.7) or (2.8)-(2.9) is of the desired type if $n = N = p$ and $\mu_i = \frac{1}{N} = \nu_j$.

2.4. **Consistency.** In our schemes to compute the surface transportation distance, for example by solving (2.8), we have so far not included any constraints on the regularity of the resulting optimal transportation plan π^*. When computing the distance between a surface and a reasonable deformation of the same surface, one does indeed find, in practice, that the minimizing π^* is fairly smooth, because neighboring points have similar neighborhoods. There is no guarantee, however, that this has to happen. Moreover, we will be interested in comparing surfaces that are far from (almost) isometric, given by noisy datasets. Under such circumstances, the minimizing π^* may well "jump around". In this subsection we propose a regularization procedure to avoid such behavior.

Computing how two surfaces best correspond makes use of the values of the "distances in similarity" $d_{\mu,\nu}^R(z_i, w_j)$ between pairs of points that "start" on one surface and "end" on the other; computing these values relies on finding a minimizing Möbius transformation for the functional (1.9). We can keep track of these minimizing Möbius transformations m_{ij} for the pairs of points (z_i, w_j) proposed for optimal correspondence by the optimal transport algorithm described above. Correspondence pairs (i,j) that truly participate in some close-to-isometry map will typically have Möbius transformations m_{ij} that are very similar. This suggests a method of filtering out possibly mismatched pairs, by retaining only the set of correspondences (i,j) that cluster together within the Möbius group.

There exist many ways to find clusters. In our applications, we gauge how far each Möbius transformation m_{ij} is from the others by computing a type of ℓ_1 variance:

$$(2.12) \qquad E_V(i,j) = \sum_{(k,\ell)} \|m_{ij} - m_{k\ell}\| \, ,$$

where the norm is the Frobenius norm (also called the Hilbert-Schmidt norm) of the 2×2 complex matrices representing the Möbius transformations, after normalizing them to have determinant one. We then use $E_V(i,j)$ as a consistency measure of the corresponding pair (i,j).

3. GENERALIZATION TO SPHERE-TYPE SURFACES

So far we have restricted ourselves to disk-type surfaces, which is somewhat limiting in practice. It is fairly straightforward to generalize the ideas presented in [19] to other types of surfaces; in this part of the paper

we show how this can be done. We choose to concentrate on the common case of sphere-type surfaces, that is, genus zero surfaces. We will start by making the necessary theoretical changes, and then we will present the numerical algorithm; an example will be given in Section 5, alongside examples for disk-type surfaces.

3.1. **Generalization of the distance function.** The Uniformization theory for sphere-type surfaces ensures a conformal one-to-one and onto mapping of the surface to the 2-sphere or equivalently to the extended complex plane $\widehat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$, the Stone-Čeck compactification of $\mathbb{C}$. The group of automorphisms of the extended plane are the Möbius transformations $m : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$, given by:

$$(3.1) \qquad\qquad m(z) = \frac{az + b}{cz + d},$$

where $a, b, c, d \in \mathbb{C}$ and $ad - bc \neq 0$. In other words, any bijective conformal mapping taking the extended plane to itself is a Möbius transformation, and vice-versa any Möbius transformation is a bijective conformal map of the extended plane.

The key to successful generalization to this case is choosing the neighborhoods $\Omega_{z_0,R}$ for a point z_0 in a Möbius-invariant way. In contrast to the situation on the disk, where the hyperbolic distance is invariant under the group $M_{\mathcal{D}}$, the extended complex plane $\widehat{\mathbb{C}}$ does not posses a distance invariant under the full Möbius group $M_{\widehat{\mathbb{C}}}$. This can be understood by noting that there is no non-constant continuous two-argument function $f(z, w)$ such that $f(m(z), m(w)) = f(z, w)$ for all Möbius transformations $m \in M_{\widehat{\mathbb{C}}}$ and all $z, w \in \widehat{\mathbb{C}}$. Therefore, the neighborhoods must be constricted in a different way.

We tackle this problem by starting with the most basic invariant of Möbius geometry, namely (generalized) circles. The neighborhood of a point z_0 will then be defined as the interior of a particular circle in the extended plane.

Let us first define the collection of *circles* in $\widehat{\mathbb{C}}$ plane with prescribed *orientation* by $\mathcal{C}$. The role of the orientation attached to circles will become clear momentarily.

Definition 3.1.

(1) A *circle* $c \in \mathcal{C}$ is defined as the set of complex numbers satisfying an equation of the type

$$\mathfrak{A} z\bar{z} + \mathfrak{B} z + \overline{\mathfrak{B} z} + \mathfrak{D} = 0,$$

where $\mathfrak{A}, \mathfrak{D} \in \mathbb{R}$ and $\mathfrak{B} \in \mathbb{C}$. Note that we define $\infty \in c$ if ∞ is an accumulation point of c (in the extended complex plane topology).

(2) The *orientation* of a circle $c \in \mathcal{C}$ is defined by labeling the "inside" and "outside" connected parts of $\widehat{\mathbb{C}} \setminus c$.

Note that this definition of a "circle" includes straight lines that can be thought of as circles through infinity (which is a legitimate point in the extended complex plane), and also "empty circles" that are empty sets (e.g. if $\mathfrak{A} = \mathfrak{D} = 1$, $\mathfrak{B} = 1$).

The candidate neighborhoods, needed to generalize the definition of $d_{\mu,\nu}^{R}(\cdot, \cdot)$, will be defined by selecting particular oriented circles in $\mathcal{C}$; since Möbius transformations already take circles to circles, we need to ensure only that our selection criterium is invariant under Möbius transformations as well. We shall of course pick the neighborhood of $z_0 \in \widehat{\mathbb{C}}$ from the collection of oriented circles $\mathcal{C}$ that contain z_0. In addition, the choice should be: 1) isometry-invariant (i.e., if $\mathcal{M}, \mathcal{N}$ are isometric sphere-type surfaces, and $z_0, w_0 \in \mathcal{D}$ are corresponding points under the isometry on their uniformizations on $\widehat{\mathbb{C}}$, then the Möbius transformation m that corresponds to that isometry should map the neighborhood $\Omega_{z_0,A}$ around z_0 to the neighborhood $\Omega_{w_0,A}$ around w_0), 2) robust to noise, and 3) characterized by a "size" parameter similar to R in the disk-type surfaces.

The key idea is to single out from the collection of circles $\mathcal{C}$ a single or discrete number of circles using the surface's metric tensor. We will outline one possible construction as an example but other construction are certainly possible.

To characterize "size" we shall use area: we consider circles such that their area (w.r.t the surface's metric) is of prescribed magnitude A, that is $\{c \mid \mathrm{vol}_{\mathcal{M}}(\bar{c}) = A\}$, where $\bar{c}$ denotes the union of the interior of the circle c (defined by its orientation) and the set c. Since we assume our surfaces have unit area, $A \in (0, 1]$.

The neighborhood $\Omega_{z_0, A}$ is then defined by

$$\Omega_{z_0, A} = \overline{\operatorname*{argmin}_{c \in \mathcal{C}_{z_0, A}} \mathrm{length}_{\mathcal{M}}(c)},$$

(3.2)
$$\text{where}$$

$$C_{z_0, A} = \left\{ c \in \mathcal{C} \mid \mathrm{vol}_{\mathcal{M}}(\bar{c}) = A \text{ and } z_0 \in \bar{c} \right\}, \text{ and}$$

where $\mathrm{length}_{\mathcal{M}}(c)$ denotes the length of the curve c based on the metric of surface $\mathcal{M}$.

A few remarks are in order. Let us consider the collection of circles $c \in \mathcal{C}$ such that $\mathrm{vol}_{\mathcal{M}}(\bar{c}) = A$. We can use the Riemann sphere model which can be thought of as the standard $S^2 \subset \mathbb{R}^3$. Each circle $c \in \mathcal{C}$ is either s standard circle on S^2 or a point or an empty set. For every point $p \in S^2$ there is a one dimensional family of circles $c_{p,t}$ defined by $c_{p,t} = \left\{ q \in S^2 \mid \langle q, p \rangle = 1 - t \right\}$, where $t \in (0, 2)$: the interior is taken to be the part that contains the point p, so that $\bar{c}_{p,t} = \left\{ q \in S^2 \mid \langle q, p \rangle \geq 1 - t \right\}$. Obviously, $v : t \mapsto \mathrm{vol}_{\mathcal{M}}(c_{p,t})$ is a monotone function, and $\lim_{t \to 0} v(t) = 0$, $\lim_{t \to 1} v(t) = 1$. Lastly, $v(t)$ is a continuous function and therefore there exists a unique value t_A (depending on p) such that $v(t_A) = A$. This means that for every point $p \in S^2$ we can find a unique circle from the concentric family $\{c_{p,t}\}_{p \in S^2, t \in (0,2)}$ that has area A. Since every non-empty and non-point circle in $\mathcal{C}$ can be identified as $c_{p,t}$ for some point $p \in S^2$ and some $t \in (0, 2)$, the collection $c_{p, t_A(p)}$, $p \in S^2$ is a parametrization of the collection $\{c \in \mathcal{C} \mid \mathrm{vol}_{\mathcal{M}}(\bar{c}) = A\}$. Now, the extra restriction $z_0 \in \bar{c}$ defines a subset of S^2 in the sense that we consider only $p \in S^2$ such that $z_0 \in \bar{c}_{p, t_A}$. Since t_A is continuous as a function of p this restriction defines a compact subset of S^2, which in turn implies that the minimum in eq.(3.2) is achieved. Lastly, on this two dimensional manifold we consider the function $c \mapsto \mathrm{length}_{\mathcal{M}}(c)$ which is a smooth function and in the generic case has a unique global minimum. We will henceforth assume that eq.(3.2) has a unique minimizer.

Since the neighborhoods $\Omega_{z_0, A}$ are chosen from the circle collection, using only intrinsic properties they are invariant to Möbius changes of coordinates of the metric; in other words if two isometric surfaces $\mathcal{M}, \mathcal{N}$ are compared at a pair of isometric points z_0, w_0 (w_0 is the image of z_0 under the isometry) then the isometry mapping $\mathcal{M}$ to $\mathcal{N}$ is a Möbius transformation taking not only z_0 to w_0 but also $\Omega_{z_0, A}$ to $\Omega_{w_0, A}$.

Once the neighborhoods are set, the definition of $d^A_{\mu, \nu}(z_0, w_0)$ is straightforward: denote by $M^A_{z_0, w_0}$ the collection of Möbius transformations m that take the interior of the circle $\Omega_{z_0, A}$ to the interior of $\Omega_{w_0, A}$, and for which $m(z_0) = w_0$. This is again a one parameter subgroup parameterized over the unit circle (angle) $\theta \in [0, 2\pi)$, as in the disk-type surface case we then define

(3.3)
$$d^A_{\mu, \nu}(z_0, w_0) := \inf_{m \in M^A_{z_0, w_0}} \int_{\Omega_{z_0, A}} \left| 1 - \frac{\nu((m(z)) \, |m'(z)|^2}{\mu(z)} \right| d\mathrm{vol}_M(z).$$

This is indeed the analog to eq. (1.9) for disk-type surfaces since both integrals can be written in the invariant form

$$\int_{\Omega_{z_0, A}} \left\| \widetilde{g}(z) - \left(m^* \widetilde{h} \right)(z) \right\|_{\widetilde{g}} d\mathrm{vol}_{\mathcal{M}}(z),$$

where $\widetilde{g}$ (resp. $\widetilde{h}$) is the push-forward metric of $\mathcal{M}$ (resp. $\mathcal{N}$) on $\widehat{\mathbb{C}}$, and the norm $\|\cdot\|_{\widetilde{g}}$ is the standard one, induced by $\widetilde{g}$: for a tensor $t_{ij} dx^i \otimes dx^j$, we denote $\widetilde{g} = \widetilde{g}_{ij} dx^i \otimes dx^j$, then $\|t\|^2_{\widetilde{g}} = t_{ij} t_{k\ell} \widetilde{g}^{ik} \widetilde{g}^{j\ell}$.

This area-based definition of the neighborhoods could also be used in the disk-type case; this would yield a unified definition for both cases. The difference between the new definition (above) and the old definition (hyperbolic geodesic disk) in the case of disk-type surfaces is that the old definition provides smaller neighborhoods near the boundary for disk type surfaces, while the new definition will maintain constant area neighborhoods even arbitrarily close to the surface's boundary; depending on the application and data properties, one or the other selection may be preferable.

3.2. **Numerical details.** The algorithm for the sphere-type case is basically the same as for the disk-type; That is, we first sample $N = n = p$ equally distributed points (as described in Section 2) Z,W on the surfaces $\mathcal{M}$,$\mathcal{N}$ (respectively). Second, the cost function $d^A_{\mu,\nu}(z_i, w_j)$ is computed between every pair of sample points $(z_i, w_j) \in Z \times W$, and finally, a discrete mass-transportation problem is solved between the discrete measures μ_Z and ν_W to output the distance and the correspondences. A few adjustments need to be made to this algorithm for the sphere-type case: 1) precomputing (approximating) the neighborhoods $\Omega_{z_i,A}$, $z_i \in Z$, and $\Omega_{w_j,A}$, $w_j \in W$, which involves more computation than for the disk-type surfaces case, 2) representing the conformal density on the extended complex plane rather than the unit disk, and approximating the local distance $d^R_{\mu,\nu}(z_i, w_j)$, and 3) calculating the optimal transport between the discrete densities. Next we describe these adjustments in more detail.

Computing the neighborhoods $\Omega_{z_i,A}$. We describe the construction of neighborhoods $\Omega_{z_i,A}$ for every sample point $z_i \in Z$ in $\mathcal{M}$. The construction in $\mathcal{N}$ is identical. We want to find the neighborhood $\Omega_{z_i,A}$ (assumed unique) based on the definition eq.(3.2). That is, $\Omega_{z_i,A}$ is the interior of a conformal circle, has surface area A, and has minimal circumference compared to all other such circles. A circle (on S^2 or equivalently on $\widehat{\mathbb{C}}$) is defined by a triplet of points z_j, z_k, z_ℓ in the usual way. Adding the orientation, each triplet provides us with two choices of conformal circle neighborhoods. In our implementation we considered all $2\binom{N}{3}$ circles generated by the sample set Z. For each such circle (endowed with orientation) we estimate the surface area inscribed in it. If it is $\epsilon-$close to the prescribed amount A we estimate its circumference on the surface. We define $\Omega_{z_i,A}$ the one with smallest circumference that contains z_i.

Approximating $d^R_{\mu,\nu}(z_i, w_j)$. The second issue that arises when generalizing to sphere-type surfaces is the representation of the conformal density $\mu(z), \nu(w)$. One option is to use a spherical interpolation scheme and repeat the steps as described in Appendix B. However, one can also pick a different path that is very simple and offers an alternative to the smooth TPS approximation described in Appendix B. The idea is to represent the conformal density by keeping track of a set of equally spread points $\widetilde{Q} = \{\widetilde{q}_\ell\}_{\ell=1}^{\mathcal{L}} \subset \mathcal{M}$ on the surface (similarly to Z), where each point represent a surface patch (Voronoi cell) of size $\frac{1}{\mathcal{L}}$. In our implementation for sphere-type surfaces, we use this latter choice. We usually take a set of size $\mathcal{L} \approx 1000$. The discrete density is then represented on the extended plane as $Q = \{q_\ell\}_{\ell=1}^{\mathcal{L}} = \Phi(\widetilde{Q}) \subset \widehat{\mathbb{C}}$. It can be shown that given a domain $\Omega \subset \widehat{\mathbb{C}}$ we have the approximation:

$$(3.4) \qquad \left| \frac{1}{\mathcal{L}} \sum_{q_i \in \Omega} 1 - \int_\Omega d\mathrm{vol}_\mathcal{M} \right| \leq C\varphi(\{\widetilde{q}_\ell\}).$$

To justify this approximation result let us denote by V_i the Voronoi cells of the set $\widetilde{Q}$ based on the metric of surface $\mathcal{M}$. Lemma D.2 (proved in Appendix D) then implies that $\frac{1}{\mathcal{L}} \sum_{q_i \in \Omega} 1 = \mathrm{vol}_\mathcal{M}(\cup V_i) \leq \mathrm{vol}_\mathcal{M}\left(\cup_{p \in \mathcal{M}} B_g(p, 2\varphi_g(\widetilde{Q}) + \epsilon) \right)$, for arbitrary $\epsilon > 0$, and it is not hard to see that for domains Ω with regular boundary curves $\left| \mathrm{vol}_\mathcal{M}\left(\cup_{p \in \mathcal{M}} B_g(p, 2\varphi_g(\widetilde{Q}) + \epsilon) \right) - \mathrm{vol}_\mathcal{M}(\Omega) \right| \leq C\varphi(\{\widetilde{q}_\ell\})$.

Note that the above arguments do not use the uniform property of $\widetilde{Q}$, and will work also when the sampling is not uniform, as long as the fill distance $\varphi_g(\widetilde{Q}) \to 0$.

Figure 1 shows the density points spread on a cat model and on its uniformization sphere, as well as the neighborhood (in red) of a single point on the cat's front leg. Denote by $P = \{p_\ell\}_{\ell=1}^{\mathcal{L}} \subset \widehat{\mathbb{C}}$ the density points for surface N. To approximate $d^A_{\mu,\nu}(z_0, w_0)$ as defined in eq.(3.3) we follow the following steps: first, we map
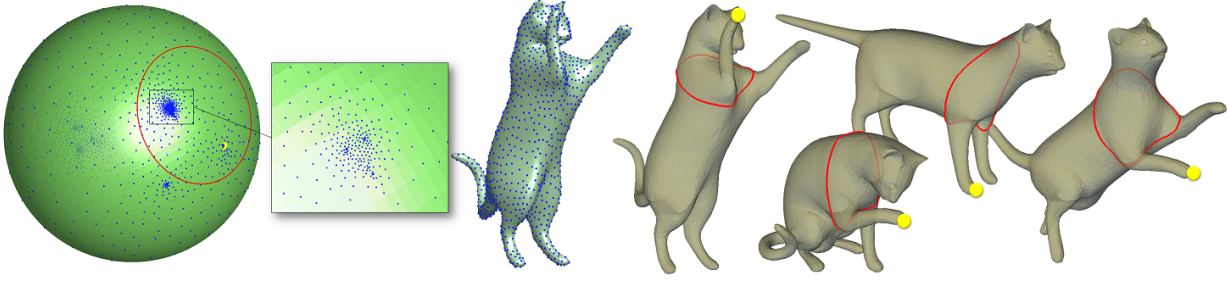
FIGURE 1. Sampling of the sphere-type cat surface (third from the left), and the sampling shown on the uniformization sphere (left). Note the zoom-in inset of the cat's head. We also show the neighborhood (marked with red on the sphere) for $A = 0.3$ of a point (marked with yellow ball on the sphere) on the cat's front leg (the model is taken from the Non-rigid World data-set [5]). On the right we show the neighborhood corresponding to the yellow point on the original cat surface model (fourth from the right) and on few other surfaces from the same class (the neighborhoods were computed independently in each surface). Note the invariance of this neighborhood under nearly-isometric deformations.

$Q \cap \Omega_{z_0,R}$ to the unit disk $\mathcal{D}$ via a Möbius transformation that is defined by taking z_0 to the origin and $\Omega_{z_0,A}$ to the unit disk $\mathcal{D}$. Denote the resulting unit disk points by Q_{z_0}. Similarly we map $P \cap \Omega_{w_0,A}$ to the unit disk (taking w_0 to the origin). We denote the resulting set by P_{w_0}. Second, for each $\theta \in \left\{0, \frac{2\pi}{L}, 2\frac{2\pi}{L}, ..., (L-1)\frac{2\pi}{L}\right\}$ we rotate Q_{z_0} by θ around the origin, $e^{\mathbf{i}\theta}Q_{z_0}$, and compare to the second density P_{w_0}. The way we compare the two densities is justified by eq. (3.4), that is we subdivide the unit disk (actually the entire square $[-1,1]^2$) into bins and count, for each of the two densities $e^{\mathbf{i}\theta}Q_{z_0}, P_{w_0}$ the number of points in each bin. Then, we can sum the absolute value of the difference to achieve our approximation to $d_{\mu,\nu}^A(z_0, w_0)$. To smoothen the approximation, it is useful to convolve the bins' structure with some kernel (this is analog to the smoothing splines used in the disk case). In our experiments (presented in Section 5) we used a 30×30 bin structure and convolved with the kernel $\frac{1}{9}(1,1,1) \otimes (1,1,1)$.

Solving the linear programming for spheres. Once we have defined $Z = \{z_i\}_{i=1}^n, W = \{w_j\}_{j=1}^p$, μ_Z, ν_W and $\check{d}_{\mu,\nu}^A(z_i, w_j)$ we can go ahead and calculate $T_{\check{d}}(\mu_Z, \nu_W)$ as explained in Section 2.3. Since our analysis in Appendix D is for general compact separable and complete metric spaces it will hold also for the sphere case. Hence, once the approximation error of $\check{d}^A(z_i, w_j) \approx d_{\mu,\nu}^A(z_i, w_j)$ is set (as outlined above), Theorem 2.4 can be applied to yield the convergence result.

4. STABILITY

In this section we prove Theorem 1.2; this is a first result connecting the new distance with local geodesic distortion. We will prove it for the disk case, but similar arguments can be used for the sphere case.

Proof. (of Theorem 1.2) We will use three different metrics on $\mathcal{D}$: the metrics of $\mathcal{M}, \mathcal{N}$ represented by the tensors $\widetilde{g}, \widetilde{h}$ (resp.), and the hyperbolic metric $d_H(\cdot, \cdot)$.

The main idea of the proof is to use that a small value for $\mathbf{d}^R(\mathcal{M}, \mathcal{N})$ means that there exists a $\pi \in \Pi(\mu, \nu)$ with respect to which the integral (1.2) of the cost function $d_{\mu,\nu}^R(z, w)$ is small as well; by the definition (1.9) of this cost function there must therefore be many corresponding neighborhoods $\Omega_{z_0,R}$ and $\Omega_{w_0,R}$ in $\mathcal{M}$ and $\mathcal{N}$ respectively that are very similar; we shall use these similarities to build local isometries.

Denote $\mathbf{d}^R(\mathcal{M}, \mathcal{N}) = \varepsilon$. Fix $K \subset \mathcal{D}$ to be a hyperbolic disk centered at the origin with an arbitrarily large (but finite) radius, $K = \Omega_{0,L}$. (Note: we could equally well have picked K to be an arbitrary set that is compact in the hyperbolic metric; this particular choice alleviates notations.) Because $\Omega_{0,L+R}$ is a compact

subset of $\mathcal{D}$, there exists a constant $C = C(L) < \infty$ such that $d_H(z, z') \leq Cd_g(z, z')$ for all $z, z' \in \Omega_{0,L+R}$. Similarly, there exist positive constants $C' = C'(L)$, $C_1 = C_1(L) < \infty$ such that $\mu(z) \geq \frac{1}{C'}$ for all $z \in \Omega_{0,L+R}$ and $\text{vol}_\mathcal{M}(\Omega_{z,R/2}) \geq \frac{2}{C_1}$ for all $z \in \Omega_{0,L} = K$.

Now set $r < R/2C$. We will prove the desired bounds for arbitrary points $z_1, z_2 \in K$ such that $d_g(z_1, z_2) < r$. Let us pick such an arbitrary pair, which we shall keep fixed for the moment. We immediately note that $d_H(z_1, z_2) < R/2$.

Let now γ_{z_1,z_2} be the minimal-length-geodesic curve connecting z_1 and z_2 (in terms of the metric corresponding to surface $\mathcal{M}$); by taking $r > 0$ sufficiently small we can ensure that this geodesic is unique. Since $d_g(z_1, \xi) < r$ and thus $d_H(z_1, \xi) < R/2$ for all $\xi \in \gamma_{z_1,z_2}$, it follows that $\gamma_{z_1,z_2} \subset \Omega_{z_1,R/2}$. Morever, by a simple application of the triangle inequality, we have $\gamma_{z_1,z_2} \subset \Omega_{z_0,R}$ for all z_0 such that $d_H(z_1, z_0) < R/2$), i.e. for all $z_0 \in \Omega_{z_1,R/2}$. It follows that

$$\Omega_{z_1,R/2} \subset B := \{z_0 \mid \gamma_{z_1,z_2} \subset \Omega_{z_0,R}\} .$$

On the other hand, $\text{vol}_\mathcal{M}(\Omega_{z_1,R/2}) \geq \inf_{z \in K} \text{vol}_\mathcal{M}(\Omega_{z,R/2}) = \min_{z \in K} \text{vol}_\mathcal{M}(\Omega_{z,R/2}) > \frac{2}{C_1}$. This implies that the volume of B on the surface $\mathcal{M}$ can be bounded from below by

$$\text{vol}_\mathcal{M}(B) \geq \text{vol}_\mathcal{M}(\Omega_{z_1,R/2}) \geq \frac{2}{C_1} .$$

We can use this lower bound to show that there must be points z_0 in B, and corresponding points w_0 in $\mathcal{D}$, for which $d^R_{\mu,\nu}(z_0, w_0)$ is small. Indeed, let be π^* an optimal transportation plan realizing the minimal transportation cost, then:

$$\varepsilon = \int_{\mathcal{D} \times \mathcal{D}} d^R_{\mu,\nu}(z, w) d\pi^*(z, w) \geq \int_{B \times \mathcal{D}} d^R_{\mu,\nu}(z, w) d\pi^*(z, w)$$

$$\geq \inf_{(z,w) \in B \times \mathcal{D}} \left[d^R_{\mu,\nu}(z, w)\right] \int_{B \times \mathcal{D}} d\pi^*(z, w) = \inf_{(z,w) \in B \times \mathcal{D}} \left[d^R_{\mu,\nu}(z, w)\right] \text{vol}_\mathcal{M}(B).$$

There thus exists some point $(z_0, w_0) \in B \times \mathcal{D}$ such that

$$d^R_{\mu,\nu}(z_0, w_0) \leq \frac{2\,\varepsilon}{\text{vol}_\mathcal{M}(B)} \leq C_1\,\varepsilon.$$

Next, we note that $d^R_{\mu,\nu}$ can be written as (see [19], and eq.(1.9))

$$d^R_{\mu,\nu}(z_0, w_0) = \inf_{m \in M_\mathcal{D}, m(z_0)=w_0} \int_{\Omega_{z_0,R}} \left\|\widetilde{g} - m^*\widetilde{h}\right\|_{\widetilde{g}(z)} d\text{vol}_\mathcal{M}(z) = \int_{\Omega_{z_0,R}} |\mu(z) - \nu(m(z))|\, d\text{vol}_H(z),$$

where $\|\cdot\|_{\widetilde{g}(z)}$ is the norm in the relevant tensor space as defined in Section 3 (and in [19]). The $m \in M_\mathcal{D}$ that satisfy $m(z_0) = w_0$ constitute a one-parameter compact family, so that the infimum is achieved; let us call f^{z_0} this minimizing Möbius transformation from $\Omega_{z_0,R}$ to $\Omega_{w_0,R}$. We have thus

$$\int_{\Omega_{z_0,R}} \left\|\widetilde{g} - (f^{z_0})^*\widetilde{h}\right\|_{\widetilde{g}(z)} d\text{vol}_\mathcal{M}(z) = \int_{\Omega_{z_0,R}} |\mu(z) - \nu(f^{z_0}(z))|\, d\text{vol}_H(z) \leq C_1\,\epsilon .$$

Now

$$s(z) = |\mu(z) - \nu(f^{z_0}(z))|$$

is Lipschitz as a function of the argument z, since μ, ν are Lipschitz on $\Omega_{0,L+2R} \subset \mathcal{D}$, and f^{z_0} is analytic; moreover, by observing that $\Omega_{z_0,R} \subset \Omega_{0,L+2R}$, we can bound the Lipschitz constant for s independently of the particular choices made so far, i.e., for all $z, z' \in \Omega_{z_0,R}$,

$$|s(z) - s(z')| \leq \kappa\,|z - z'| .$$

Take now any u in $\Omega_{z_0,R}$, and set $S = s(u)$. Then

$$C_1\,\epsilon \geq \int_{\Omega_{z_0,R}} \max(0, S - \kappa|u - z|)\, d\text{vol}_H(z) \geq C''\,S^3 ,$$

where $C'' > 0$ can be picked independently of the location of u within $\Omega_{z_0,R}$, and uniformly for $z_0 \in \Omega_{0,L+R}$. It follows that $S \leq C_2 \, \epsilon^{1/3}$. This shows that

$$(4.1) \qquad \max_{z \in \Omega_{z_0,R}} \left\| \widetilde{g} - m^* \widetilde{h} \right\|_{\widetilde{g}(z)} \leq C_2 \, \varepsilon^{1/3} \,,$$

for some constant C_2 that doesn't depend on ϵ or z_1, z_2.

Given arbitrary z_1, z_2 in K, we have thus found $\Omega_{z_0,R}$ that contains the full geodesic γ_{z_1,z_2} and a Möbius map f^{z_0} from $\Omega_{z_0,R}$ to a corresponding $\Omega_{w_0,R}$ that, within to a small error controlled by the small quantity $\mathbf{d}^R(\mathcal{M},\mathcal{N})$, maps the local geometry in $\mathcal{M}$ to that in $\mathcal{N}$.

To alleviate notations in what follows, we drop the superscript z_0 on f^{z_0}. Our plan is to use the minimizing geodesic path $\gamma_{z_1,z_2}(t) : [0,1] \to \mathcal{M}$ between the points $\gamma_{z_1,z_2}(0) = z_1$, and $\gamma_{z_1,z_2}(1) = z_2$ to compute bounds on $d_h(f(z_1), f(z_2))$. Using the differential $[Df]$ of the map f, which is a linear map between the tangent spaces $T_z \mathcal{M}$ (with the metric $\widetilde{g}$) and $T_w \mathcal{N}$ (with the metric $\widetilde{h}$), we have

$$d_h(f(z_1), f(z_2)) \leq \int_0^1 \left\| \frac{d}{dt} f(\gamma_{z_1,z_2}(t)) \right\|_h dt$$

$$\leq \int_0^1 \|[Df]\| \, \|\dot{\gamma}_{z_1,z_2}(t)\|_g \, dt$$

$$(4.2) \qquad\qquad \leq \max_{z \in \gamma_{z_1,z_2}} \|[Df(z)]\| \, d_g(z_1, z_2) \,,$$

which shows that an upper bound on $\max_{z \in \gamma_{z_1,z_2}} \|[Df(z)]\|$ will give us one of the desired inequalities.

The second inequality is achieved using that

$$d_g(f^{-1}(f(z_1)), f^{-1}(f(z_2))) = \int_0^1 \left\| \frac{d}{dt} f^{-1}(\gamma_{f(z_1),f(z_2)}(t)) \right\|_g dt$$

$$\leq \int_0^1 \left\| [Df^{-1}] \right\| \, \left\| \dot{\gamma}_{f(z_1),f(z_2)}(t) \right\|_h dt$$

$$(4.3) \qquad\qquad \leq \max_{w \in \gamma_{f(z_1),f(z_2)}} \left\| [Df(w)]^{-1} \right\| d_h(f(z_1), f(z_2)).$$

To use this, we thus need to upper bound $\max_{w \in \gamma_{f(z_1),f(z_2)}} \left\| [Df(w)]^{-1} \right\|$.

In the remainder of this proof, we show how bounds on $\max_{z \in \gamma_{z_1,z_2}} \|[Df(z)]\|$ and $\max_{w \in \gamma_{f(z_1),f(z_2)}} \left\| [Df(w)]^{-1} \right\|$ can be derived from (4.1).

First, we take an orthonormal basis $E = \{e_1, e_2\} \subset T_z \mathcal{M}$. That is,

$$(4.4) \qquad\qquad e_k^i e_\ell^j g_{ij} = e_k^i e_\ell^j \widetilde{\mu}(z) \delta_{i,j} = \delta_{k,\ell}.$$

Similarly we take orthonormal basis $B = \{b_1, b_2\} \subset T_w \mathcal{N}$.

We will denote the matrix $[Df] = [Df_z]$ representing the differential Df_z of f at the point z, in the bases E, B. The norm of Df_z is the induced norm

$$\|Df_z\| = \max_{\xi \in T_z \mathcal{M}, \xi \neq 0} \frac{\|Df(\xi)\|_{\widetilde{h}(f(z))}}{\|\xi\|_{\widetilde{g}(z)}}.$$

Writing the tensor $\widetilde{g}$ in the basis E we get the Euclidean form $\widetilde{g}(z) = de_1 \otimes de_1 + de_2 \otimes de_2$, and the tensor $\widetilde{h}$ in the basis B will have the same form $\widetilde{h}(w) = db_1 \otimes db_1 + db_1 \otimes db_2$. The pull-back $f^* \widetilde{h}(z)$ will have the form $f^* \widetilde{h}(z) = ([Df_z]^t [Df_z])_{ij} \, de_i \otimes de_j$. Therefore in the basis E we have

$$\left\| \widetilde{g} - f^* \widetilde{h} \right\|_{\widetilde{g}(z)} = \left\| Id - [Df]^t [Df] \right\|_F,$$

where Id is the 2×2 identity matrix, and $\|\cdot\|_F$ denotes the Frobenius norm.

Writing the singular value decomposition of $[Df]$ we have
$$[Df] = Q \operatorname{diag}(\sigma_1(z), \sigma_2(z)) R^T,$$
where Q, R are orthogonal and $\sigma_1 \leq \sigma_2$ are the respective singular values. Then we have
(4.5)
$$\|\widetilde{\mu} - f^*\widetilde{\nu}\|_{g(z)}^2 = \left\| Id - [Df]^T[Df] \right\|_F^2 = \left\| R \left(Id - \begin{pmatrix} (\sigma_1(z))^2 & 0 \\ 0 & (\sigma_2(z))^2 \end{pmatrix} \right) R^T \right\|_F^2 = (1-\sigma_1(z)^2)^2 + (1-\sigma_2(z)^2)^2,$$

Using (4.1),(4.5) we get
$$\left| 1 - \sigma_2(z)^2 \right| \leq C_2 \varepsilon^{1/3}.$$
From this last bound on $\sigma_2(z)$ for $z \in \Omega_{z_0, R}$ there exists a constant $C_3 > 0$ such that
$$0 \leq \sigma_2(z) \leq 1 + C_3 \varepsilon^{1/3}.$$
Since $\|[Df]\| = \sigma_2$, this is the desired estimate for the first inequality.

For the second inequality, we need to bound $\left\|[Df]^{-1}\right\| = \sigma_1^{-1}$. The computation (4.5) shows that
$$\left| 1 - \sigma_1(z)^2 \right| \leq C_2 \varepsilon^{1/3},$$
from which we obtain
$$0 \leq \frac{1}{\sigma_1(z)} \leq 1 + C_4 \varepsilon^{1/3},$$
which concludes our argument.

$\square$

5. EXPERIMENTAL VALIDATION AND COMMENTS

In this section we perform experimental validation of our algorithms. We have tested and experimented with our algorithms on four different data-sets:

(1) *Non-rigid World data-set* [5]. This data-set was distributed by Bronstein,Bronstein and Kimmel and was specifically constructed for evaluating shape comparison algorithms in the scenario of non-rigid shapes; it contains meshes of different objects (cats, dogs, wolves, humans, etc...) in different poses. We compare our results on this data-set to the Gromov-Hausdorff algorithm suggested in [1, 5].
(2) *SHREC 2007 Watertight Benchmark* [10] This data-set contains meshes of several objects within a given *semantic* class for several different classes, such as chairs, 4-legged animals, humans, etc... It is more challenging for isometric-invariant matching algorithms since most of the objects are far from isometric to the objects in the same semantic class, for example the 4-legged animals class contains a giraffe and a dog.
(3) *Synthetic.* We constructed this data-set to test the effect of the "size" parameter R on the distance behavior.
(4) *Primate molar teeth.* This data-set originates from a real biological problem/application; it consists of molar teeth surface for different primates. It was communicated to us by biologists who compare these shapes for characterization and classification of mammals.

Remark 5.1. For all data-sets, we scaled the meshes to have unit area, because our goal is to compare surfaces solely based on shape, regardless of size.

Non-rigid World data-set and comparison to Gromov-Hausdorff-type distance. In the first experiment we ran our sphere-type algorithm to determine conformal Wasserstein distances for all pairs in the Non-rigid World data-set, distributed by Bronstein, Bronstein and Kimmel [1, 5, 4]; we compare the results to those obtained using the code for the (symmetrized) partial embedding Gromov-Hausdorff (speGH) distance distributed by the same authors. The speGH distance has been used with great success in surface comparison [1, 5, 4] and can handle situations beyond the scope of our, more limited algorithm, it can e.g.,

compare surfaces of different genus. We therefore consider it as a state-of-the-art algorithm. In order to compare our Conformal Wasserstein distance with speGH for applications of interest to us, we use both algorithms on Non-rigid World data-set where all the surfaces are scaled to have unit area and where 100 sample points are chosen on each surface.

The Non-rigid World data-set contains meshes of different poses of the following articulated objects: a centaur, a cat, a dog, a horse, a human female, and two human males ("Michael" and "David").

The two resulting dissimilarity matrices are shown in Figure 2 (a,d). The dissimilarity matrices are both normalized by translating the minimal value to zero and scaling the maximal value to one. The color scheme is Matlab's "Jet" where dark blue represents low (close to zero) value, and dark red indicates high (close to one) values. The Sphere-type algorithm used $A = 0.3$, and a 30×30 bins discretization with the convolution kernel $(1, 1, 1) \otimes (1, 1, 1)$ (all the sphere-type examples use these bin settings) to obtain the discretizations of the conformal density. The timing for running one comparison $\mathbf{d}^A(\mathcal{M}, \mathcal{N})$ was 90 seconds on 2.2GHz AMD Opteron processor. Figure 2 (b-c) and (e-f) shows two nearest neighbors classifications tests where a white square inside a dark-red area means success and white square on black area is a failure.

The structure of the dissimilarity matrix is illustrated in the two plots in Figure 3. Figure 3 (a) shows the classification rates as a function of the number K of nearest neighbors, where for each fixed K we calculated the classification rate as follows. For each object we counted how many from its K-nearest neighbors are of the same class. We summed all these numbers and divided by the total number of possible correct classifications. The blue curve shows the analysis of the dissimilarity matrix output by our distance algorithm and the red curve by the speGH distance code. In (b) we show the ROC curve where for every $K = 1, 2, .., 10$ we plot the True Positive Rate (TPR), that is the number of true positives divided by the number of positives, as a function of the False Positive Rate (FPR), that is the number of false positives divided by the number of negatives.

SHREC 2007 Watertight Benchmark [10]**.** Our next experiment deals with a data-set with larger in-class variations; SHREC 2007 contains 20 categories of models with 20 meshes for each category (400 meshes in total) the categories are, e.g., chairs, 4-legged animals, humans, planes, tables, etc...

For our experiment, we restricted ourselves to all the meshes of 8 categories that contained only surfaces of genus zero (since our current algorithm does not support surfaces of higher genus) and that seemed intrinsically similar; these categories were: humans, 4-legged animals, ants, hands, airplanes, teddy-bears, pliers, Armadillos. We ran our sphere-type algorithm to compute the distance between all pairs. We tested the "size" parameters $A = 0.3, 0.4, 0.5$. The bin was the same as for the previous data-set. However, to achieve faster running times (we had about 25,000 comparisons...) we took only 50 sample points. The running time for one pair of objects was around 15 seconds.

Figure 4 shows the dissimilarity matrices (top row) using the three different values if A: $0.3, 0.4, 0.5$, and the dissimilarity matrix resulting from combining them:

$$T_d(\mu, \nu) = T_d^{0.3}(\mu, \nu) \cdot T_d^{0.4}(\mu, \nu) \cdot T_d^{0.5}(\mu, \nu).$$

Note that $T_d(\mu, \nu)$ is also a metric and suggests a way to remove the influence of the size parameter if desired. The combined distance produced the best classification results as seen in the bottom row, where for each row the white square shows the nearest neighbor to that object. The dark red areas represent the different categories. The combined distance reached the very high classification rate of 95% on this challenging data-set. Figure 5 shows for one object in each category its four closest neighbors. Note the non-rigid nature of some of the objects (e.g., humans, hands), and the substantial deviation from perfect isometry within class (e.g., 4-legged animals). Figure 6 demonstrates a partial failure case where although the first two nearest neighbors to the Giraffe are within the 4-legged animals category, the third nearest neighbor is the one-armed armadillo, which belong to a different category (remember that the algorithm is size invariant).

Synthetic data-set. This experiment was designed to test the influence of the size parameter R on the behavior of the distance. The surfaces we compared are shown in the top row of Figure 7, they each have three

(a) This paper's dissimilarity matrix

(b) 1st nearest neighbor
94% classification rate

(c) 5 nearest neighbors
87% classification rate

(d) speGH dissimilarity matrix

(e) 1st nearest neighbor
78% classification rate

(f) 5 nearest neighbors
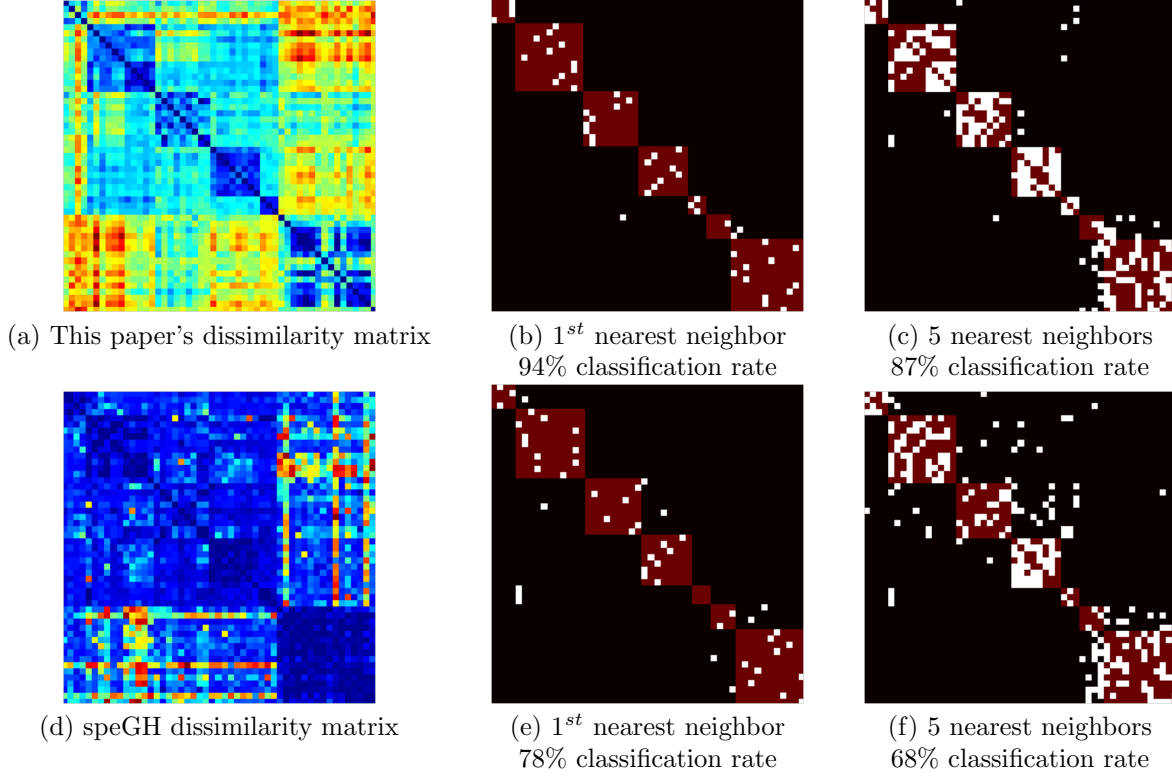68% classification rate

FIGURE 2. The dissimilarity matrices for the unit area scaled Non-rigid World data-set [1, 5] calculated with the conformal Wasserstein (CW) distance given by our algorithm (a) and the symmetric partial-embedding Gromov-Hausdorff (speGH) distance [1, 5] (d). In the second column (b,e) we show the ground truth classification matrix (dark-red) and the first nearest neighbor (white) according to the CW (in b) and to the speGH (in e) distance for each row. The third column (c,f) shows the five nearest neighbors (when there are fewer than five in some category we simply limit ourselves to the number in that category). Note that white squares should be in dark red regions to indicate correct classification. Note that with only 100 sampling points, "Michael" and "David" are not distinguishable, so if a pose of Michael is among the nearest neighbors of a pose of David, we still count it as a correct classification.

small bumps, in different positions. At first sight, one might think that for small R, the distance $\mathbf{d}^R(\mathcal{M}, \mathcal{N})$ based on comparing neighborhoods of "size" R, would have trouble distinguishing these objects from each other. However, one should keep in mind that the uniformization process is a global one: changing the metric in one region of the surface would effect the uniformization of other regions (but influence would decay like appropriate Green's function). Figure 7 plot the distance of disk-type model A to the four others, for different R values. We also show hyperbolic neighborhoods corresponding to the three different size parameters (color coded the same way as the graphs). We scaled the distances to have maximum of one (since smaller R results naturally in smaller distances). Note that even the smallest size value $R = 0.25$ already distinguishes between the different models. Further note, that larger size parameters such as $R = 0.75$ results in slightly more intuitive linear distance behavior. Overall, the size parameter R does affect the distance, but not in a very significant way.

Primate molar teeth. Finally, we present a few experimental results related to a biological application; in a case study of the use of our approach to the characterization of mammals by the surfaces of their molars, we compare high resolution scans of the masticating surfaces of molars of several lemurs, which are

(a) Correct Classification Rate
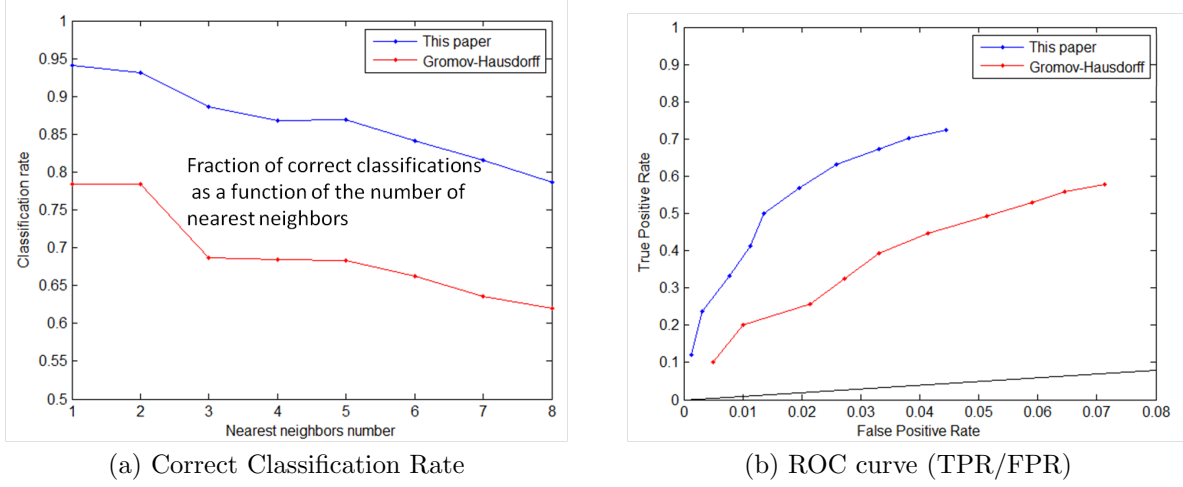
(b) ROC curve (TPR/FPR)

FIGURE 3. Correct Classification Rate and ROC curves for the dissimilarity matrices produces by our method (blue) and Gromov-Hausdorff-type metric (red). See text for details.
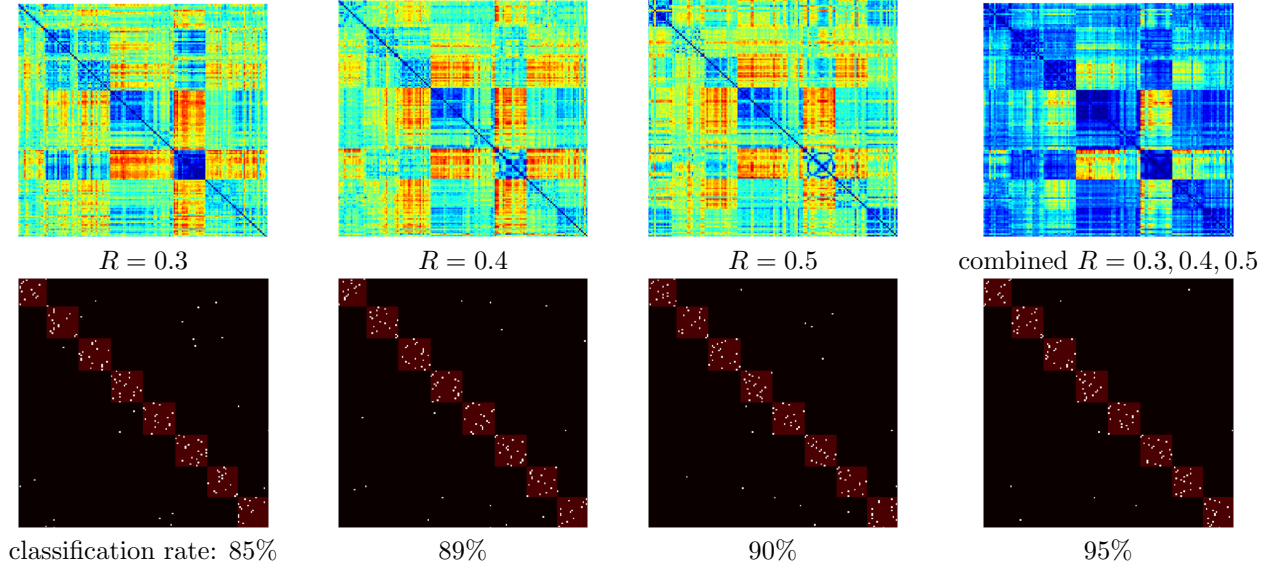

$R = 0.3$ $R = 0.4$ $R = 0.5$ combined $R = 0.3, 0.4, 0.5$

classification rate: 85%   89%   90%   95%

FIGURE 4. Dissimilarity matrices (top row) for the SHREC 2007 watertight Benchmark [10] with different size parameter: $R = 0.3, 0.4, 0.5$ and their combination (see text for details). Bottom row shows the first nearest neighbor classification test where white squares denote the nearest neighbor of that row's object and the dark red area represent correct category classification.

small primates living in Madagascar. Traditionally, biologists specializing in this area carefully determine landmarks on the tooth surfaces, and measure characteristic distances and angles involving these landmarks. A first stage of comparing different tooth surfaces is to identify correspondences between landmarks. Figure 8 illustrates how $d_{\mu,\nu}^R(z, w)$ (disk-type) can be used to find corresponding pairs of points on two surfaces by showing both a "good" and a "bad" corresponding pair. The left two columns of the figure show the pair of points in each case; the two middle columns show the best fit after applying the minimizing Möbius on the corresponding disk representations; the rightmost column plots $\int_{\Omega_{z_0,R}} |\mu(z) - (m_{z_0,w_0,\sigma}^* \nu)(z)| \, d\mathrm{vol}_H(z)$, the value of the "error", as a function of parameter σ, parameterizing the Möbius transformations that map
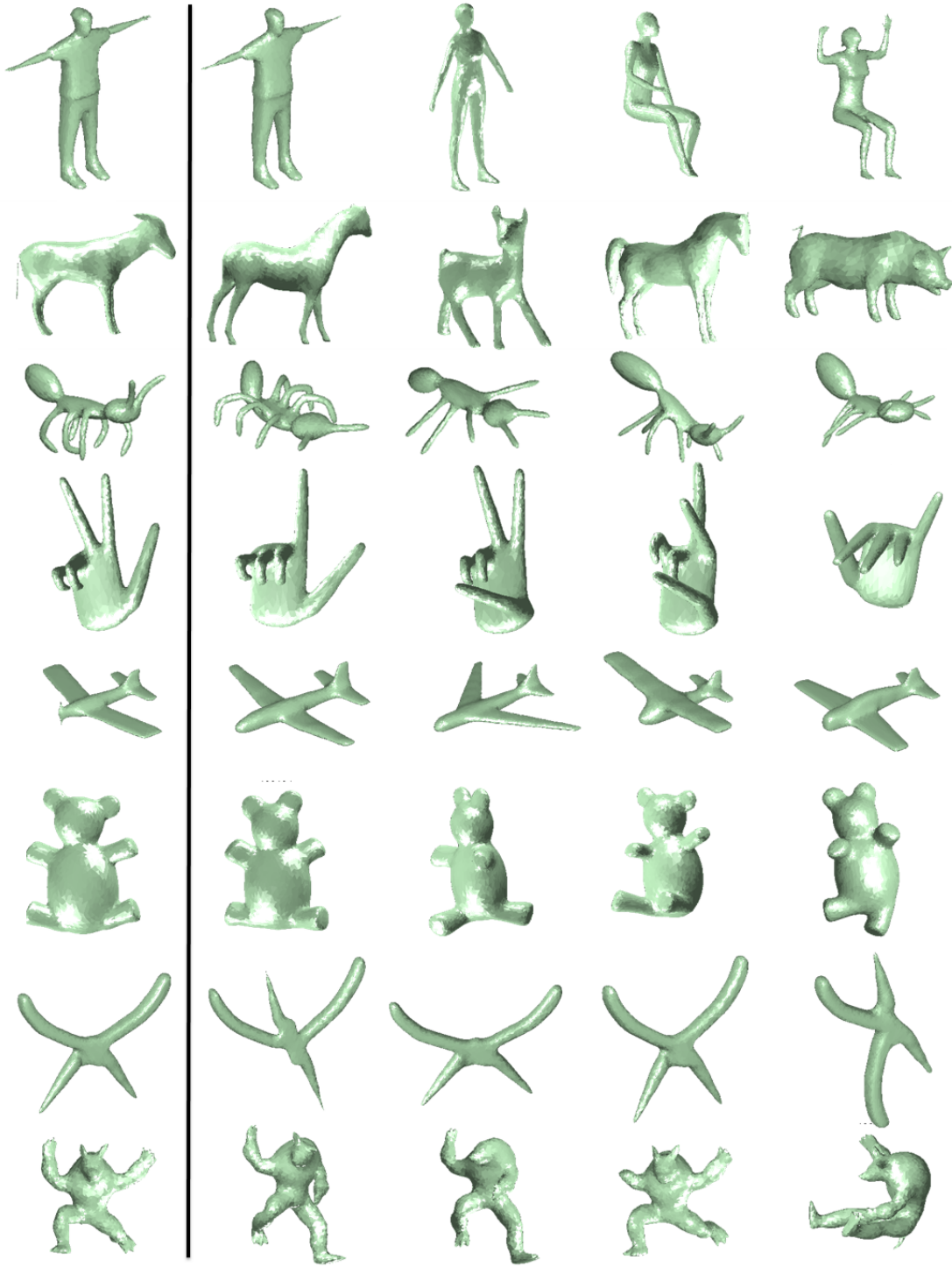
Figure 5. SHREC 2007 watertight Benchmark [10]: we show the four closest neighbors to each of the objects on the left side (we show one example from each category).

a given point z_0 to another given point w_0 (they are parameterized over S^1, see Lemma 3.5 in [19]). The
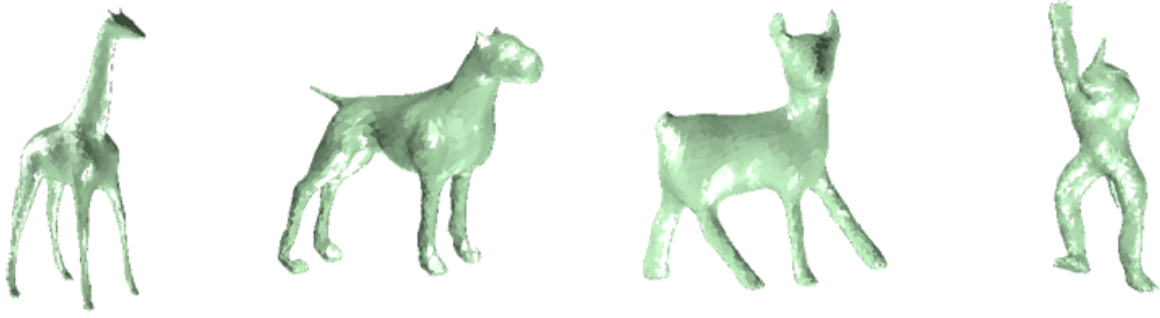
FIGURE 6. SHREC 2007 watertight Benchmark [10]: a partially failure case where the giraffe has correct two nearest neighbors, however its 3rd nearest neighbor is a one armed armadillo. Remember that our matching is scale invariant.
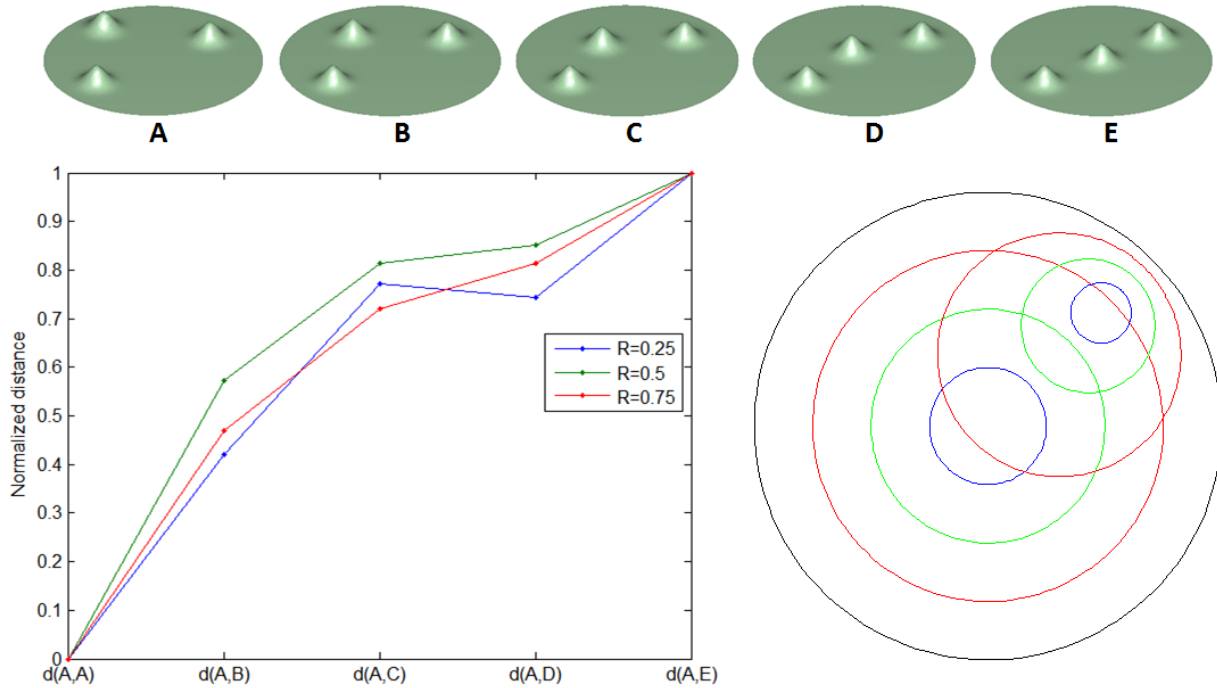
FIGURE 7. Testing influence of the size parameter R to sperate identical models with small features. We compare the distances of the disk model marked with A to all other models $A - E$ using three different size parameters $R = 0.25, 0.5, 0.75$. We also show examples of hyperbolic disk of these radii as used by our disk-type algorithm (the are color coded like the graph lines).

"best" corresponding point w_0 for a given z_0 is the one that produces the lowest minimal value for the error, i.e. the lowest $d_{\mu,\nu}^R(z_0, w_0)$.

Figure 9 show the top 120 most consistent corresponding pairs (in groups of 20) for two molars belonging to lemurs of different species. Corresponding pairs are indicated by highlighted points of the same color.
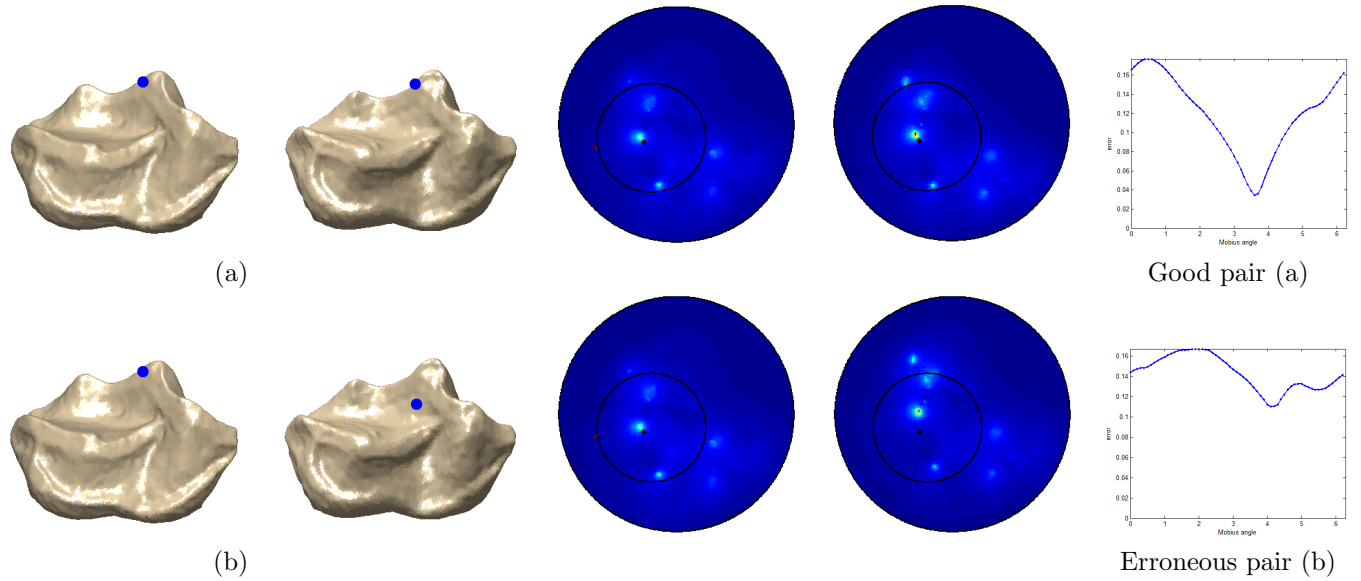
FIGURE 8. Calculation of the local distance $d^R_{\mu,\nu}(\cdot,\cdot)$ between pairs of points on two different surfaces (each row shows a different pair of points; the two surfaces are the same in the top and bottom rows). The first row shows a "good" pair of points together with the alignment of the conformal densities $\mu, m^*\nu$ based on the best Möbius transformation m minimizing $d^R_{\mu,\nu}$. The plot of this latter integral as a function of m (parameterized by $\sigma \in [0, 2\pi)$, see (1.5)) is shown in the right-most column. The second row shows a "bad" correspondence which indeed leads to a higher local distance $d^R_{\mu,\nu}$.

These correspondences have surprised the biologists from whom we obtained the data sets; their experimental measuring work, which incorporates finely balanced judgment calls, had defied earlier automatization attempts.

Once the differences and similarities between molars from different animals have been quantified, they can be used (as part of an approach) to classify the different individuals. Figure 10 illustrates a preliminary result that illustrates the possibility of such classifications based on the distance operator between surfaces introduced in this paper. The figure illustrates the pairwise distance matrix for eight molars, coming from individuals in four different species (indicated by color). The clustering was based on only the distances between the molar surfaces; it clearly agrees with the clustering by species, as communicated to us by the biologists from whom we obtained the data sets.

One final comment regarding the computational complexity of our method. There are two main parts: the preparation of the distance matrix d_{ij} and the linear programming optimization. For the linear programming part we used a Matlab interior point implementation with N^2 unknowns, where N is the number of points spread on the surfaces. In our experiments, the optimization typically terminated after $15 - 20$ iterations for $N = 150 - 200$ points, which took about 2-3 seconds. The computation of the similarity distance d_{ij} took longer, and was the bottleneck in our experiments. We separate the disk-type and the sphere-type algorithms.

For the sphere-type algorithm if we use $N = \mathcal{L}$ sample points (Q) on each surface (see Section 3) then for each pair we compare the difference (using fixed size bin structure) of the discrete conformal densities for fixed number of Möbius transformations. This results in $O(N^3)$ algorithm for computing the distance matrix d_{ij}. In our experiments the total distance computation time (including linear programming optimization) was around 15 seconds for $N = \mathcal{L} = 50$ (in the SHREC 2007 data-set), to 90 second per comparison for

$N = \mathcal{L} = 100$ (in the Non-rigid world data-set). In the sphere-type examples we have used 2.2GHz AMD Opteron processor. The sphere-type algorithm was coded completely in Matlab and was not optimized.

For the disk-type algorithm, if we spread N points on each surface, and use them all to interpolate the conformal factors Γ_μ, Γ_ν, if we use P points in the integration rule, and take L points in the Möbius discretization (see Section 2 for details) then each approximation of $d_{\mu,\nu}^R(z_i, w_j)$ by (2.5) requires $O(L \cdot P \cdot N)$ calculations, as each evaluation of Γ_μ, Γ_ν (Thin-Plate Spline approximation we use to interpolate the conformal densities, described in Appendix B) takes $O(N)$ and we need $L \cdot P$ of those. Since we have $O(N^2)$ distances to compute, the computation complexity for calculating the similarity distance matrix d_{ij} is $O(L \cdot P \cdot N^3)$. This step was coded in C++ (and therefore the time difference to the sphere-type case) and took 3.5 seconds for $N = 50$, 51 seconds for $N = 100$, under 5 minutes for $N = 150$ and two hours for $N = 300$ (in these examples we took $P \approx N$). However, also in this case we have not optimized the algorithm and we believe these times can be reduced significantly. The disk-type algorithm ran on Intel Xeon (X5650) 2.67GHz processor.

6. ACKNOWLEDGMENTS

REFERENCES

1. R. Kimmel A. M. Bronstein, M. M. Bronstein, *Generalized multidimensional scaling: a framework for isometry-invariant partial surface matching*, Proc. National Academy of Sciences (PNAS) **103** (2006), no. 5, 1168–1172.

2. Mikael Fortelius Jukka Jernvall Alistair R. Evans, Gregory P. Wilson, *High-level similarity of dentitions in carnivorans and rodents*, Nature **445** (2007), 78–81.

3. Susanne C. Brenner and L. Ridgway Scott, *The mathematical theory of finite element methods*, third ed., Texts in applied mathematics, vol. 15, 2008.

4. Alexander Bronstein, Michael Bronstein, and Ron Kimmel, *Numerical geometry of non-rigid shapes*, 1 ed., Springer Publishing Company, Incorporated, 2008.

5. Alexander M. Bronstein, Michael M. Bronstein, and Ron Kimmel, *Efficient computation of isometry-invariant distances between surfaces*, SIAM J. Sci. Comput. **28** (2006), no. 5, 1812–1836.

6. E. Cela, *The quadratic assignment problem: Theory and algorithms (combinatorial optimization)*, Springer, 1998.

7. Gerhard Dziuk, *Finite elements for the Beltrami operator on arbitrary surfaces*, vol. 1357, Springer Berlin / Heidelberg, 1988.

8. Y. Eldar, M. Lindenbaum, M. Porat, and Y. Zeevi, *The farthest point strategy for progressive image sampling*, 1997.

9. Bruce Fischl, Martin I. Sereno, Roger B. H. Tootell, and Anders M. Dale, *High-resolution intersubject averaging and a coordinate system for the cortical surface*, Hum. Brain Mapp **8** (1999), 272–284.

10. Daniela Giorgi, Silvia Biasotti, and Laura Paraboschi, *SHREC:SHape REtrieval Contest: Watertight models track, http://watertight.ge.imati.cnr.it/*, 2007.

11. Mikhail Gromov, M. Katz, P. Pansu, and S. Semmes, *Metric structures for riemannian and non-riemannian spaces*, Birkhäuser Boston, December 2006.

12. Xianfeng Gu and Shing-Tung Yau, *Global conformal surface parameterization*, SGP '03: Proceedings of the 2003 Eurographics/ACM SIGGRAPH symposium on Geometry processing (Aire-la-Ville, Switzerland, Switzerland), Eurographics Association, 2003, pp. 127–137.

13. Steven Haker, Lei Zhu, Allen Tannenbaum, and Sigurd Angenent, *Optimal mass transport for registration and warping*, International Journal on Computer Vision **60** (2004), 225–240.

14. Irwin Kra Hershel M. Farkas, *Riemann surfaces*, Springer, 1992.

15. Hildebrandt, Klaus, Polthier, Konrad, Wardetzky, and Max, *On the convergence of metric and geometric properties of polyhedral surfaces*, Geometriae Dedicata **123** (2006), no. 1, 89–112.

16. L. Kantorovich, *On the translocation of masses*, C.R. (Dokl.) Acad. Sci. URSS (N.S.) **37** (1942), 199–201.

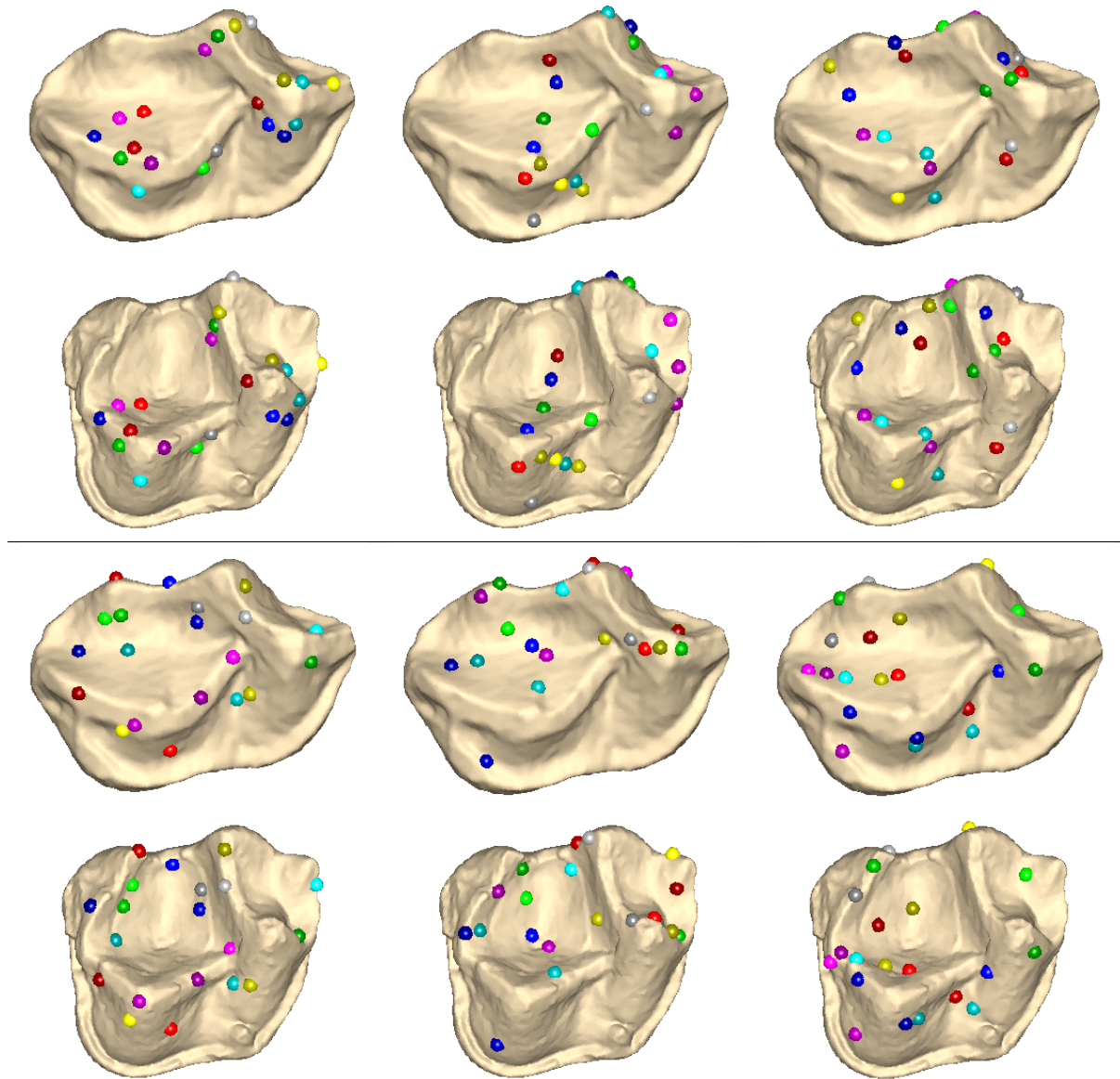17. M.D. Plummer L. Lovász, *Matching theory*, North-Holland, 1986.

FIGURE 9. The top 120 most consistent corresponding pairs between two molar teeth models.

18. Parsons LM Liotti M Freitas CS Rainey L Kochunov PV Nickerson D Mikiten SA Fox PT Lancaster JL, Woldorff MG, *Automated talairach atlas labels for functional brain mapping*, Human Brain Mapping **10** (2000), 120–131.

19. Yaron Lipman and Ingrid Daubechies, *Conformal wasserstein distances: Comparing surfaces in polynomial time*, Advances in Mathematics (ELS), accepted for publication. (2011).

20. Yaron Lipman and Thomas Funkhouser, *Mobius voting for surface correspondence*, ACM Transactions on Graphics (Proc. SIGGRAPH) **28** (2009), no. 3.

21. Facundo Memoli, *On the use of gromov-hausdorff distances for shape comparison*, Symposium on Point Based Graphics (2007).

22. Facundo Mémoli and Guillermo Sapiro, *A theoretical and computational framework for isometry invariant recognition of point cloud data*, Found. Comput. Math. **5** (2005), no. 3, 313–347.

23. Billingsley Patrick, *Convergence of probability measures*, John Wiley & Sons, 1968.

24. Ulrich Pinkall and Konrad Polthier, *Computing discrete minimal surfaces and their conjugates*, Experimental Mathematics **2** (1993), 15–36.

25. Konrad Polthier, *Conjugate harmonic maps and minimal surfaces*, Preprint No. 446, TU-Berlin, SFB 288 (2000).
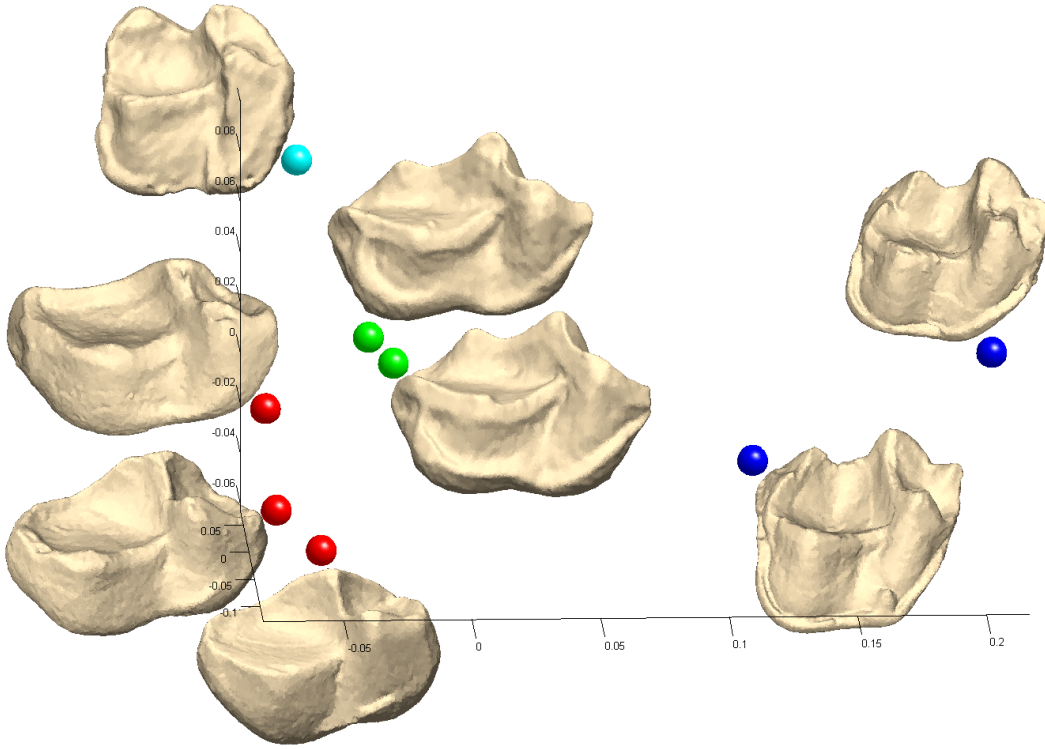
FIGURE 10. Embedding of the distance graph of eight teeth models using multi-dimensional scaling. Different colors represent different lemur species. The graph suggests that the geometry of the teeth might suffice to classify species.

26. _____, *Computational aspects of discrete minimal surfaces*, Global Theory of Minimal Surfaces, Proc. of the Clay Mathematics Institute 2001 Summer School, David Hoffman (Ed.), CMI/AMS (2005).

27. Y. Rubner, C. Tomasi, and L. J. Guibas, *The earth mover's distance as a metric for image retrieval*, International Journal of Computer Vision **40** (2000), no. 2, 99–121.

28. Conroy B Bryan RE Ramadge PJ Haxby JV Sabuncu MR, Singer BD, *Function-based intersubject alignment of human cortical anatomy*, Cereb Cortex. (2009).

29. Alexander Schrijver, *A course in combinatorial optimization, course note*, 2008.

30. Boris Springborn, Peter Schröder, and Ulrich Pinkall, *Conformal equivalence of triangle meshes*, ACM SIGGRAPH 2008 papers (New York, NY, USA), SIGGRAPH '08, ACM, 2008, pp. 77:1–77:11.

31. George Springer, *Introduction to riemann surfaces*, AMS Chelsea Publishing, 1981.

32. Cedric Villani, *Topics in optimal transportation (graduate studies in mathematics, vol. 58)*, American Mathematical Society, March 2003.

33. W. Zeng, X. Yin, Y. Zeng, Y. Lai, X. Gu, and D. Samaras, *3d face matching and registration based on hyperbolic ricci flow*, CVPR Workshop on 3D Face Processing (2008), 1–8.

34. W. Zeng, Y. Zeng, Y. Wang, X. Yin, X. Gu, and D. Samaras, *3d non-rigid surface matching and registration based on holomorphic differentials*, The 10th European Conference on Computer Vision (ECCV) (2008).

APPENDIX A.

We prove Theorem 2.1. We start with a simple lemma showing that all Möbius transformations restricted to $\Omega_{0,T}$, $T < \infty$, are Lipschitz with a universal constant, for which we provide an upper bound.

Lemma A.1. *A Möbius transformation $m \in M_\mathcal{D}$ restricted to $\Omega_{0,T}$, $T < \infty$ is Lipschitz continuous with Lipschitz constant $C_m \leq \frac{1-|a|^2}{(1-r_T|a|)^2}$, where $a = m^{-1}(0)$ and $r_T = tanh(T)$.*

Proof. Denote $m(z) = e^{\mathbf{i}\theta}\frac{z-a}{1-z\bar{a}}$. Then, for $z, w \in \Omega_{0,T}$ we have

$$\left| m(z) - m(w) \right| \le \left| e^{\mathbf{i}\theta}\frac{z-a}{1-z\bar{a}} - e^{\mathbf{i}\theta}\frac{w-a}{1-w\bar{a}} \right| \le \left| \frac{(z-a)(1-w\bar{a}) - (w-a)(1-z\bar{a})}{(1-z\bar{a})(1-w\bar{a})} \right|$$

$$\le \left| \frac{(z-w)(1-|a|^2)}{(1-z\bar{a})(1-w\bar{a})} \right| \le |z-w|\frac{1-|a|^2}{(1-r_T|a|)^2}.$$

$\square$

Next we prove:

Theorem 2.1 *For Lipschitz continuous μ, ν,*

$$\left| d^R_{\mu,\nu}(z_i, w_j) - \min_{m(z_i)=w_j}\sum_k \alpha_k \left| \mu(\widetilde{m}_i(p_k)) - \nu(m(\widetilde{m}_i(p_k))) \right| \right| \le C\,\varphi_E(\{p_k\})\ ,$$

where the constant C depends only on μ, ν, R.

Proof. First, denote $f(z) = \left| \mu(\widetilde{m}_i(z)) - \nu(m(\widetilde{m}_i(z))) \right|$. Then,

$$(A.1) \qquad \left| \int_{\Omega_{0,R}} f(z)d\mathrm{vol}_H(z) - \min_{m(z_i)=w_j}\sum_k \alpha_k f(p_k) \right| \le \sum_k \int_{\Delta_k} \left| f(z) - f(p_k) \right| d\mathrm{vol}_H(z)$$

$$\le \omega_f^{\Omega_{0,R}}(\varphi_E(\{p_k\}))\int_{\Omega_{0,R}} d\mathrm{vol}_H,$$

where Δ_k are the intersections of $\Omega_{0,R}$ with the Euclidean Voronoi cells defined by the centers p_k, and the modulus of continuity $\omega_f^{\Omega_{0,R}}(h) = \sup_{|z-w|<h;z,w\in\Omega_{0,R}} |f(z) - f(w)|$ is used. Note that

$$(A.2) \qquad\qquad \omega_f^{\Omega_{0,R}} \le \omega_{\mu\circ\widetilde{m}_i}^{\Omega_{0,R}} + \omega_{\nu\circ m\circ\widetilde{m}_i}^{\Omega_{0,R}}.$$

Denote the Lipschitz constants of μ, ν by C_μ, C_ν, respectively. From Lemma A.1 we see that, for $z, w \in \Omega_{0,R}$,

$$\left| \mu(\widetilde{m}_i(z)) - \mu(\widetilde{m}_i(w)) \right| \le C_\mu \left| \widetilde{m}_i(z) - \widetilde{m}_i(w) \right| \le C_\mu \frac{1-|a|^2}{(1-r_R|a|)^2}|z-w| \le C_\mu\frac{1}{(1-r_R)^2}|z-w|\,,$$

which is independent of $\widetilde{m}_i$. Similarly,

$$\left| \nu(m(\widetilde{m}_i(z))) - \nu(m(\widetilde{m}_i(w))) \right| \le C_\nu \left| m(\widetilde{m}_i(z)) - m(\widetilde{m}_i(w)) \right| \le C_\nu\frac{1}{(1-r_R)^2}|z-w|\,,$$

which is independent of $m, \widetilde{m}_i$. Combining these with eq. (A.1-A.2) we get

$$\left| \int_{\Omega_{0,R}} f(z)d\mathrm{vol}_H(z) - \min_{m(z_i)=w_j}\sum_k \alpha_k f(p_k) \right| \le (C_\mu + C_\nu)\frac{\int_{\Omega_{0,R}} d\mathrm{vol}_H}{(1-r_R)^2}\varphi_E(\{p_k\}),$$

which finishes the proof. $\square$

Finally, we prove:

Theorem 2.2 *For Lipschitz continuous μ, ν,*

$$\left| d^R_{\mu,\nu}(z_i, w_j) - \breve{d}^R_{\mu,\nu}(z_i, w_j) \right| \le C_1\,\varphi_E(\{p_k\}) + C_2\,L^{-1}\ ,$$

where the constants C_1, C_2 depend only on μ, ν, R.

Proof. In view of Theorem 2.1 it is sufficient to prove that

$$\left| \widehat{d}^R_{\mu,\nu}(z_i, w_j) - \breve{d}^R_{\mu,\nu}(z_i, w_j) \right| \le C\,L^{-1},$$

for an appropriate constant C depending only upon μ, ν, R. Denote by $m_{i,j}$ the minimizer of $\widehat{d}^R_{\mu,\nu}(z_i, w_j)$. Then

$$
\left| \widehat{d}^r_{\mu,\nu}(z_i, w_j) - \breve{d}^r_{\mu,\nu}(z_i, w_j) \right| = \min_{\ell=1..L} \left| \sum_k \alpha_k \Big[|\mu(\widetilde{m}_i(p_k)) - \nu(m_{i,j}(\widetilde{m}_i(p_k)))| - |\mu(\widetilde{m}_i(p_k)) - \nu(m_{z_i,w_j,2\pi\ell/L}(\widetilde{m}_i(p_k)))| \Big] \right|
$$

$$
\leq \min_{\ell=1..L} \sum_k \alpha_k \left| \nu(m_{i,j}(\widetilde{m}_i(p_k))) - \nu(m_{z_i,w_j,2\pi\ell/L}(\widetilde{m}_i(p_k))) \right|
$$

$$
\leq C_\nu \min_{\ell=1..L} \sum_k \alpha_k \left| m_{i,j}(\widetilde{m}_i(p_k)) - m_{z_i,w_j,2\pi\ell/L}(\widetilde{m}_i(p_k)) \right|
$$

$$
\leq \frac{C_\nu}{(1-r_R)^2} \min_{\ell=1..L} \sum_k \alpha_k \left| \widetilde{m}_j^{-1}(m_{i,j}(\widetilde{m}_i(p_k))) - \widetilde{m}_j^{-1}(m_{z_i,w_j,2\pi\ell/L}(\widetilde{m}_i(p_k))) \right|
$$

where, as in previous theorem, we denote by C_ν the Lipschitz constant of ν in $\mathcal{D}$, and in the last inequality we have used Lemma A.1 while taking $\widetilde{m}_j^{-1}$ as defined in Section 2. From eq.(2.4) we have that

$$
\left| \widehat{d}^r_{\mu,\nu}(z_i, w_j) - \breve{d}^r_{\mu,\nu}(z_i, w_j) \right| \leq \frac{C_\nu}{(1-r_R)^2} \min_{\ell=1..L} \sum_k \alpha_k \left| \widetilde{m}_j^{-1}(m_{i,j}(\widetilde{m}_i(p_k))) - e^{\mathbf{i}2\pi\ell/L} p_k \right|.
$$

Now note that $\widetilde{m}_j^{-1} \circ m_{i,j} \circ \widetilde{m}_i \in M_{\mathcal{D}}$ also fixes the origin; it follows that $\widetilde{m}_j^{-1} \circ m_{i,j} \circ \widetilde{m}_i(z) = e^{\mathbf{i}\theta} z$ for some $\theta \in [0, 2\pi)$. We therefore have

$$
\left| \widehat{d}^r_{\mu,\nu}(z_i, w_j) - \breve{d}^r_{\mu,\nu}(z_i, w_j) \right| \leq \frac{C_\nu}{(1-r_R)^2} \min_{\ell=1..L} \sum_k \alpha_k \left| e^{\mathbf{i}\theta} p_k - e^{\mathbf{i}2\pi\ell/L} p_k \right|
$$

$$
\leq \frac{r_R C_\nu}{(1-r_R)^2} \min_{\ell=1..L} \sum_k \alpha_k \left| e^{\mathbf{i}\theta} - e^{\mathbf{i}2\pi\ell/L} \right|
$$

$$
\leq \frac{r_R C_\nu}{(1-r_R)^2} \left(\sum_k \alpha_k \right) \min_{\ell=1..L} \left| e^{\mathbf{i}\theta} - e^{\mathbf{i}2\pi\ell/L} \right|
$$

$$
\leq \frac{r_R C_\nu \, 2\pi}{L(1-r_R)^2} \int_{\Omega_{0,R}} d\mathrm{vol}_H.
$$

$\square$

APPENDIX B.

In this appendix we review a few basic notions such as the representation of (approximations to) surfaces by faceted, piecewise flat approximations, called *meshes*, and discrete conformal mappings; the conventions we describe here are the same as adopted in [20].

We denote a triangular mesh by the triple $M = (V, E, F)$, where $V = \{v_i\}_{i=1}^m \subset \mathbb{R}^3$ is the set of vertices, $E = \{e_{i,j}\}$ the set of edges, and $F = \{f_{i,j,k}\}$ the set of faces (oriented $i \to j \to k$). When dealing with a second surface, we shall denote its mesh by N. In this appendix, we assume our mesh is homeomorphic to a disk.

Next, we introduce "conformal mappings" of a mesh to the unit disk. Natural candidates for discrete conformal mappings are not immediately obvious. In particular, it is not possible to use a continuous piecewise-affine map the restriction of which to each triangle would be a (positively oriented) similarity transformation: continuity would force the similarity transformations of any two adjacent triangles to coincide, meaning such a map would be globally a similarity. A different approach uses the notion of discrete harmonic and discrete conjugate harmonic functions due to Pinkall and Polthier [24, 26] to define a discrete conformal mapping on the mid-edge mesh. The mid-edge mesh $\mathsf{M} = (\mathsf{V}, \mathsf{E}, \mathsf{F})$ of a given mesh $M = (V, E, F)$ is defined as follows. For the vertices $\mathsf{v}_r \in \mathsf{V}$, we pick the mid-points of the edges of the mesh M; we call these the mid-edge points of M. There is thus a $\mathsf{v}_r \in \mathsf{V}$ corresponding to each edge $e_{i,j} \in E$. If v_s and v_r are the mid-points

of edges in E that share a vertex in M, then there is an edge $\mathbf{e}_{s,r} \in \mathbf{E}$ that connects them. It follows that for each face $f_{i,j,k} \in F$ we can define a corresponding face $\mathbf{f}_{r,s,t} \in \mathbf{F}$, the vertices of which are the mid-edge points of (the edges of) $f_{i,j,k}$; this face has the same orientation as $f_{i,j,k}$. Note that the mid-edge mesh is not a manifold mesh, as illustrated by the mid-edge mesh in Figure 11, shown together with its "parent" mesh: in $\mathbf{M}$ each edge "belongs" to only one face $\mathbf{F}$, as opposed to a manifold mesh, in which most edges (the edges on the boundary are exceptions) function as a hinge between two faces. This "lace" structure makes a mid-edge mesh more flexible: it turns out that it is possible to define a piecewise linear map that makes each face in $\mathbf{F}$ undergo a pure scaling (i.e. all its edges are shrunk or extended by the same factor) and that simultaneously flattens the whole mid-edge mesh (we provide more details on this flattening below). By extending this back to the original mesh, we thus obtain a map from each triangular face to a similar triangle in the plane; these individual similarities can be "knitted together" through the mid-edge points, which continue to coincide (unlike most of the vertices of the original triangles).

We have thus relaxed the problem, and we define a map via a similarity on each triangle, with continuity for the complete map at only *one* point of each edge, namely the mid point. This procedure was also used in [20]; for additional implementation details we refer the interested reader (or programmer) to that paper, which includes a pseudo-code.

Discrete mesh   Mid-edge mesh

Surface mesh zoom-in   Mid-edge mesh zoom-in
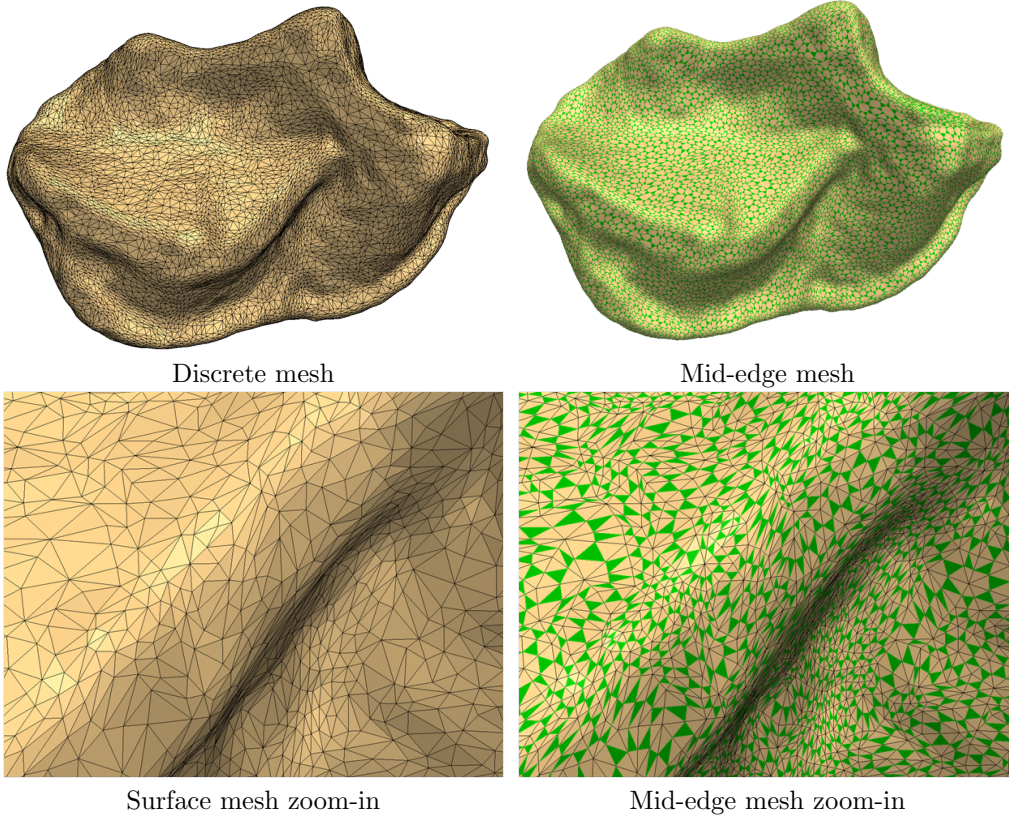
FIGURE 11. A mammalian tooth surface mesh, with the corresponding mid-edge mesh. In the mid-edge mesh, the faces are the smaller green triangles within the faces of the surface mesh.

This flattening procedure maps the boundary of the mesh onto a region with a straight horizontal slit (see Figure 12, where the boundary points are marked in red) [20]. We can assume, without loss of generality, that this slit coincides with the interval $[-2,2] \subset \mathbb{C}$. Now applying the inverse of the holomorphic map $z = w + \frac{1}{w}$ maps $\mathbb{C} \setminus [-2,2]$ conformally to the disk $\mathcal{D}$, with the slit at $[-2,2]$ mapped to the boundary

of the disk. It follows that when this map is applied to our flattened mid-edge mesh, its image is a mid-edge mesh in the unit disk, with the boundary of the disk corresponding to the boundary of our (disk-like) surface. (See Figure 12.) We shall denote by $\Phi : \mathbf{V} \to \mathbb{C}$ the composition of these different conformal and discrete-conformal maps, from the original mid-edge mesh to the corresponding mid-edge mesh in the unit disk.

Mid-edge uniformization

Uniformization Zoom-in

After mapping to the disk

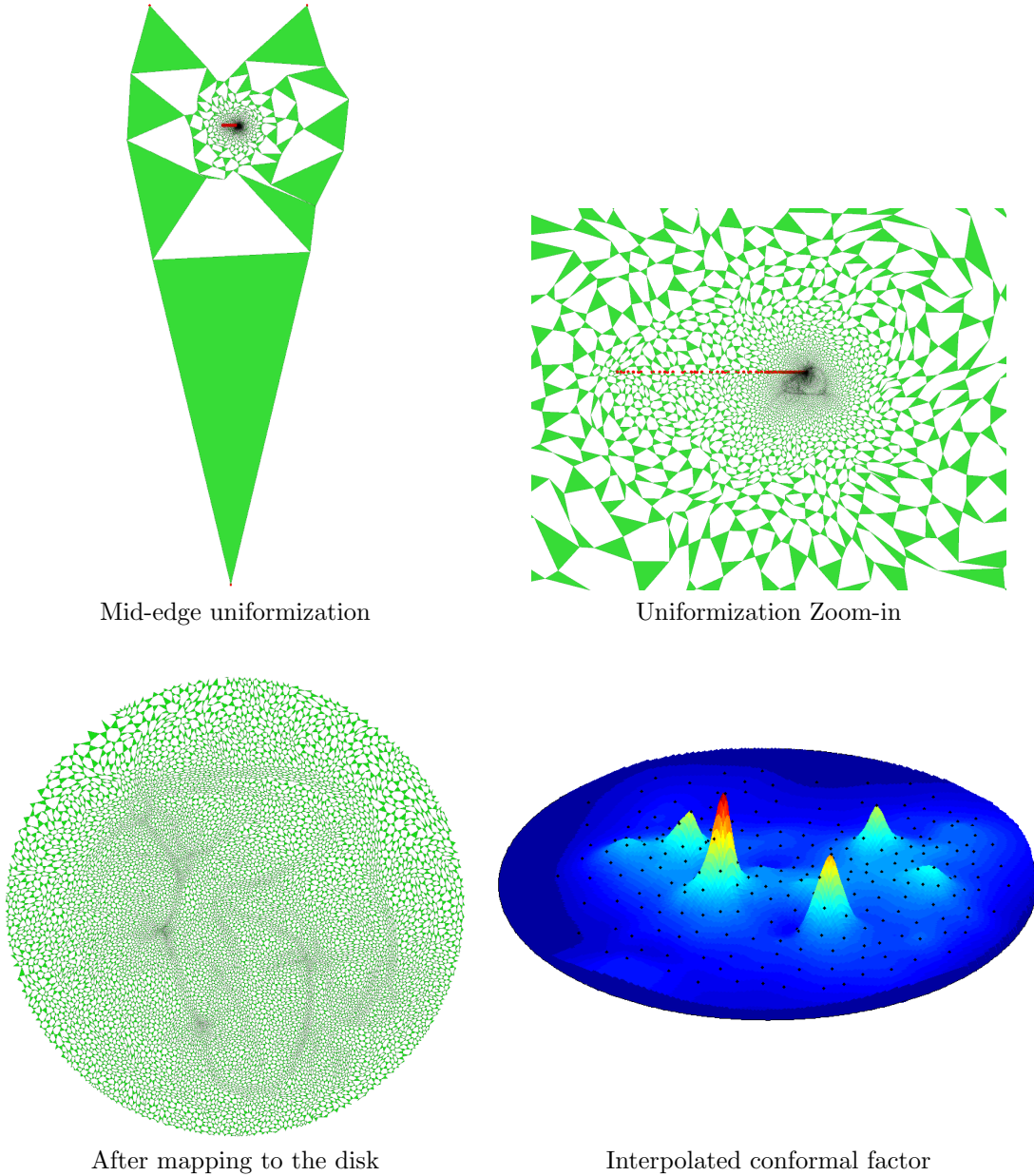Interpolated conformal factor

FIGURE 12. The discrete conformal transform to the unit disk for the surface of Figure 11, and the interpolation of the corresponding discrete conformal factors (plotted with the JET color map in Matlab). The red points in the top row's images show the boundary points of the surface.

Next, we define the Euclidean discrete conformal factors, defined as the density, w.r.t. the Euclidean metric, of the mid-edge triangles (faces), i.e.

$$\mu^E_{\mathbf{f}_{r,s,t}} = \frac{\mathrm{vol}_{\mathbb{R}^3}(\mathbf{f}_{r,s,t})}{\mathrm{vol}(\Phi(\mathbf{f}_{r,s,t}))}.$$

Note that according to this definition, we have

$$\int_{\Phi(\mathbf{f}_{r,s,t})} \mu^E_{\mathbf{f}_{r,s,t}} \, d\mathrm{vol}_E = \frac{\mathrm{vol}_{\mathbb{R}^3}(\mathbf{f}_{r,s,t})}{\mathrm{vol}_E(\Phi(\mathbf{f}_{r,s,t}))} \, \mathrm{vol}_E(\Phi(\mathbf{f}_{r,s,t})) = \mathrm{vol}_{\mathbb{R}^3}(\mathbf{f}_{r,s,t}),$$

where vol_E denotes the standard Lebesgue (Euclidean) volume element in $\mathcal{D}$, and $\mathrm{vol}_{\mathbb{R}^3}(\mathbf{f})$ stands for the area of $\mathbf{f}$ as induced by the standard Euclidean volume element in $\mathbb{R}^3$. The discrete Euclidean conformal factor at a mid-edge vertex $\mathbf{v}_r$ is then defined as the average of the conformal factors for the two faces $\mathbf{f}_{r,s,t}$ and $\mathbf{f}_{r,s',t'}$ that touch in $\mathbf{v}_r$, i.e.

$$\mu^E_{\mathbf{v}_r} = \frac{1}{2}\left(\mu^E_{\mathbf{f}_{r,s,t}} + \mu^E_{\mathbf{f}_{r,s',t'}}\right).$$

Figure 12 illustrates the values of the Euclidean conformal factor for the mammalian tooth surface of earlier figures. The discrete hyperbolic conformal factors are defined according to the following equation, consistent with the convention adopted in section 1,

(B.1) $$\mu^H_{\mathbf{v}_r} = \mu^E_{\mathbf{v}_r}\left(1 - |\Phi(\mathbf{v}_r)|^2\right)^2.$$

As before, we shall often drop the superscript H: unless otherwise stated, $\mu = \mu^H$, and $\nu = \nu^H$.

The (approximately) conformal mapping of the original mesh to the disk is completed by constructing a smooth interpolant $\Gamma_\mu : \mathcal{D} \to \mathbb{R}$, that interpolates the discrete conformal factor so far defined only at the vertices in $\Phi(\mathbf{V})$; Γ_ν is constructed in the same way. In practice we use Thin-Plate Splines, i.e. functions of the type

(B.2) $$\Gamma_\mu(z) = p_1(z) + \sum_i b_i \, \psi(|z - z_i|),$$

where $\psi(r) = r^2 \log(r^2)$, $p_1(z)$ is a linear polynomial in x^1, x^2, and $b_i \in \mathbb{C}$; p_1 and the b_i are determined by the data that need to be interpolated. Similarly $\Gamma_\nu(w) = q_1(w) + \sum_j c_j \, \psi(|w - w_j|)$ for some constants $c_j \in \mathbb{C}$ and a linear polynomial $q_1(w)$ in y^1, y^2. We use as interpolation centers two point sets $Z = \{z_i\}_{i=1}^n$, and $W = \{w_j\}_{j=1}^p$ that are uniformly distributed over the surfaces $\mathcal{M}$ and $\mathcal{N}$ (resp.), they are (relatively small) subsets of the mid-edge mesh vertex sets. In practice we calculate these (sub) sample sets by starting from an initial random seed on the surface (which will itself not be included in the set), and take the geodesic furthest point in $\mathbf{V}$ (we approximate geodesic distances with Dijkstra's algorithm) from the seed as the initial point of the sample set. One then keeps repeating this procedure, selecting at each iteration the point that lies at the furthest geodesic distance from the set of points already selected. This algorithm is known as the Farthest Point Algorithm (FPS) [8]. An example of the output of this algorithm, using geodesic distances on a disk-type surface, is shown in Figure 13. Further discussion of practical aspects of Voronoi sampling of a surface can be found in [4].
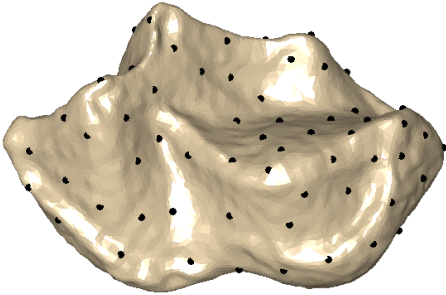

FIGURE 13. Sampling of the surface of Figure 11 obtained by the Farthest Point Algorithm.

Once the sample sets Z and W are determined we project them to the uniformization space using Φ. The bottom-right part of Figure 12 shows the result of the interpolation based on the centers Z, W (shown as black points).

To compute the explicit Thin-Plane-Splines (B.2), we use a standard smoothing Thin-Plate Spline procedure:

$$\Gamma_\mu(z) = \operatorname*{argmin}_\gamma \left\{ \lambda \sum_r |\mu_{\mathbf{v}_r} - \gamma(\Phi(\mathbf{v}_r))|^2 + (1-\lambda) \int_{\mathcal{D}} \left(\frac{\partial^2 \gamma}{(\partial x^1)^2} \right)^2 + \left(\frac{\partial^2 \gamma}{\partial x^1 \partial x^2} \right)^2 + \left(\frac{\partial^2 \gamma}{(\partial x^2)^2} \right)^2 dx^1 \wedge dx^2 \right\},$$

where the minimization is over all γ in the appropriate Sobolev space and where we picked the values $0.95 - 0.99$ manually (it was fixed per dataset of surfaces) for the smoothing factor λ to avoid over-fitting the data. We noticed that λ does not have large effect on the results.

In our implementation we assumed we have a smooth representation of the conformal factors $\mu(z) = \Gamma_\mu(z)$, $\nu(w) = \Gamma_\nu(w)$ and we simply use the notation μ, ν for these approximations.

To conclude this whirlwind description of the algorithm and ideas we use for discrete uniformization, we provide a short exposition on discrete and conjugate discrete harmonic functions on triangular meshes as in [7, 24, 25, 26], and show how we use them to conformally flatten disk-type (or even just simply connected) triangular meshes.

Discrete harmonic functions are defined using a variational principle in the space of continuous piecewise linear functions defined over the mesh $PL_{\mathcal{M}}$ ([7]), as follows. Let us denote by $\phi_i(z)$, $i = 1, .., m$, the scalar functions that satisfy $\phi_j(v_i) = \delta_{i,j}$ and are affine on each triangle $f_{i,j,k} \in F$. Then, the (linear) space of continuous piecewise-linear function on M can be written in this basis:

$$PL_M = \left\{ \sum_{i=1}^m u_i \phi_i(z) \mid (u_1, ..., u_m)^T \in \mathbb{R}^m \right\}.$$

Next, the following quadratic form is defined over PL_M:

$$(\text{B.3}) \qquad E_{Dir}(u) = \sum_{f \in F} \int_f \langle \nabla u, \nabla u \rangle \, d\mathrm{vol}_{\mathbb{R}^3},$$

where $\langle \cdot \rangle = \langle \cdot \rangle_{\mathbb{R}^3}$ denotes the inner-product induced by the ambient Euclidean space, and $d\mathrm{vol}_{\mathbb{R}^3}$ is the induced volume element on f. This quadratic functional, the *Dirichlet energy*, can be written as follows:

$$(\text{B.4}) \qquad E_{Dir}\left(\sum_i u_i \phi_i \right) = \sum_{i,j=1}^m u_i u_j \left[\sum_{f \in F} \int_f \langle \nabla \phi_i, \nabla \phi_j \rangle \right] d\mathrm{vol}_{\mathbb{R}^3} = \sum_{i,j=1}^m u_i u_j \int_M \langle \nabla \phi_i, \nabla \phi_j \rangle \, d\mathrm{vol}_{\mathbb{R}^3}.$$

The discrete harmonic functions are then defined as the functions $u \in PL_M$ that are critical for $E_{Dir}(u)$, subject to some constraints on the boundary of M. The linear equations for discrete harmonic function $u \in PL_M$ are derived by partial derivatives of E_{Dir}, (B.4) w.r.t. $u_i, i = 1, .., m$:

$$(\text{B.5}) \quad \frac{\partial E_{Dir}(u)}{\partial u_k} = 2 \sum_{i=1}^m u_i \left[\sum_{f \in F} \int_f \langle \nabla \phi_i, \nabla \phi_k \rangle \right] d\mathrm{vol}_{\mathbb{R}^3} = 2 \int_M \langle \nabla u, \nabla \phi_k \rangle \, d\mathrm{vol}_{\mathbb{R}^3} = 2 \int_{R_k} \langle \nabla u, \nabla \phi_k \rangle \, d\mathrm{vol}_{\mathbb{R}^3},$$

where $R_k \subset M$ is the 1-ring neighborhood of vertex v_k. The last equality uses that ϕ_k is supported on R_k.

Now, let $u = \sum_i u_i \phi_i$ be a discrete harmonic function. Pinkall and Polthier observed that conjugating the piecewise-constant gradient field ∇u (constant on each triangle $f \in F$), i.e. rotating the gradient ∇u in each triangle f by $\pi/2$ in the positive (= counterclockwise) sense (we assume M is orientable), results in a new vector field $*du = Jdu$ with the special property that its integrals along (closed) paths that cross edges only at their mid-points are systematically zero (see for example [26]). This means in particular that we can define a piecewise linear function $*u$ such that its gradient satisfies $d * u = *du$ and that is furthermore continuous through the mid-edges $\mathbf{v} \in \mathbf{V}$. The space of piecewise-linear functions on meshes that are continuous through

the mid-edges is well-known in the finite-element literature, where it is called $ncPL_M$, the space of non-conforming finite elements [3]. The Dirichlet form (B.3) is defined over the space of non-conforming elements $ncPL_M$ as well; the non-conforming discrete harmonic functions are defined to be the functions $v \in ncPL_M$ that are critical for E_{Dir} and that satisfy some constraints on the mid-edges of the boundary of the mesh. Polthier [26] shows that if $u \in PL_M$ is a discrete harmonic function, then $*u \in ncPL_M$ is also discrete harmonic, with the same Dirichlet energy, and vice-versa. Solving for the discrete harmonic function after fixing values at the boundaries amounts to solving a sparse linear system which is explicitly given in [26].

This theory can be used to define discrete conformal mappings, and used to flatten a mesh in a "discrete conformal" manner, as follows. The flattening is done by constructing a pair of conjugate piecewise linear functions $(u, *u)$ where $u \in PL_M$, $*u \in ncPL_M$, and the flattening map $\Phi : \mathbf{M} \to \mathbb{C}$ is given by

$$(B.6) \qquad\qquad\qquad\qquad \Phi = u + \mathbf{i} * u.$$

Since $d * u = J du$, Φ is a similarity transformation on each triangle $f \in F$. Furthermore, Φ is continuous through the mid-edge points $\mathbf{v}_r \in \mathbf{V}$; Φ is thus well-defined on the points in $\mathbf{V}$ and maps them to the complex plane.

The function u is defined by choosing an arbitrary triangle $f_{out} \in F$, excising it from the mesh, setting the values of u at two of f_{out}'s vertices u_{i_1}, u_{i_2} to 0 and 1, respectively, and then solving for the discrete harmonic u that satisfies these constraints. See for example Figure 12 (top-left); the "missing mid-edge face" corresponding to the excised face f_{out} would have connected the three mid-edge vertices that have a only one mid-edge face touching them. The conjugate function $*u$ is constructed by a simple conjugation (and integration) process as described in [26] and [20].

A surprising property of the Discrete Uniformization Φ as it is defined above, which nicely imitates the continuous theory (see [31]) is that it takes the boundaries of $\mathcal{M}$ to horizontal slits, see Figure 12, top row (boundary vertices colored in red). This property allows us to easily construct a closed form analytic map (with "analytic" in its standard complex analytic sense) that will further bijectively map the entire complex plane $\mathbb{C}$ minus the slit to the open unit disk, completing our Uniformization procedure.

The proof of this property is similar to that for Proposition 35 in [26]; see also [20]. More precisely:

Theorem B.1. *Let $\Phi : \mathbf{M} \to \mathbb{C}$ be the flattening map from the mid-edge mesh $\mathbf{M}$ of a mesh $\mathcal{M}$ with boundary, using a discrete harmonic and conjugate harmonic pair as described above. Then, for each connected component of the boundary of $\mathcal{M}$, the mid-edge vertices of boundary edges are all mapped onto one line segment parallel to the real axis.*

Proof. Suppose $u = \sum_i u_i \phi_i(\cdot)$ is a discrete harmonic, piecewise linear and continuous function, defined at each vertex $v_i \in V$, excluding the two vertices of the excised triangle for which values are prescribed; then we have, by (B.5),

$$(B.7) \qquad\qquad\qquad\qquad \int_{R_i} \langle \nabla \phi_i, \nabla u \rangle d\mathrm{vol}_{\mathbb{R}^3} = 0,$$

Next, consider a boundary vertex v_j of the mesh $\mathcal{M}$. Denote by $\mathbf{v}_r, \mathbf{v}_s$ the mid-edge vertices on the two boundary edges touching vertex v_j. We will show that $*u(\mathbf{v}_r) = *u(\mathbf{v}_s)$; this will imply the theorem, since $*u$ gives the imaginary coordinate for the images of the mid-edge vertices under the flattening map (see (B.6)) . Observe that on the triangle $f_{i,j,k}$,

$$(B.8) \qquad\qquad\qquad\qquad \nabla \phi_j = \frac{J(v_i - v_k)}{2 \operatorname{vol}_{\mathbb{R}^3}(f_{i,j,k})}.$$

Recalling that $\nabla * u = J\nabla u$, using (B.8), and $J^T = -J$, we obtain

$$*u(\mathbf{v}_r) - *u(\mathbf{v}_s) = \int_\gamma d * u = \int_\gamma *du = \sum_{f_{i,j,k} \ni v_j} \left\langle J\nabla u \mid_{f_{i,j,k}}, \frac{1}{2}(v_i - v_k) \right\rangle$$

$$= \sum_{f_{i,j,k} \ni v_j} \left\langle \nabla u \mid_{f_{i,j,k}}, \frac{1}{2} J^T (v_i - v_k) \right\rangle = \sum_{f_{i,j,k} \ni v_j} \left\langle \nabla u \mid_{f_{i,j,k}}, -\nabla \phi_j \mid_f \right\rangle \mathrm{vol}_{\mathbb{R}^3}(f)$$

$$= -\int_M \langle \nabla u, \nabla \phi_j \rangle \, d\mathrm{vol}_{\mathbb{R}^3} = 0,$$

where γ is the piecewise linear path starting at $\mathbf{v}_r$ and passing through the mid-edge vertices of the 1-ring neighborhood of v_j ending at $\mathbf{v}_s$. The last equality is due to (B.5). □

A natural question, when dealing with any type of finite-element approximation, concerns convergence as the mesh is refined: convergence in what sense, and at what rate? For discrete harmonic functions over meshes, this convergence is discussed in [15, 25]. These convergence results are in the weak sense; this motivated our defining the discrete conformal factors $\mu_{\mathbf{f}}$ via integrated quantities (volumes) in Section 2.

Finally, we note that the method presented here for Discrete Uniformization is just one option among several; other authors have suggested other techniques; for example [12]. Typically, this part of the complete algorithm described in this paper could be viewed as a "black box": the remainder of the algorithm would not change if one method of Discrete Uniformization is replaced by another.

APPENDIX C.

In this Appendix we prove a lemma used in the proof of Theorem 2.6.

Lemma *The $N \times N$ matrices π satisfying*

(C.1)
$$\begin{cases} \sum_i \pi_{ij} \leq 1 \\ \sum_j \pi_{ij} \leq 1 \\ \pi_{ij} \geq 0 \\ \sum_{i,j} \pi_{ij} = M < N \end{cases}$$

constitute a convex polytope $\mathcal{P}$ of which the extremal points are exactly those π that satisfy all these constraints, and that have all entries equal to either 0 or 1.

Remark. Note that the matrices $\pi \in \mathcal{P}$ with all entries in $\{0, 1\}$ have exactly M entries equal to 1, and all other entries equal to zero; if one removes from these matrices all rows and columns that consist of only zeros, what remains is a $M \times M$ permutation matrix.

Proof. $\mathcal{P}$ can be considered as a subset of $\mathbb{R}^{N^2}$, with all entries nonnegative, summing to M. The two inequalities in (C.1) imply that the entries of any $\pi \in \mathcal{P}$ are bounded by 1. These inequalities can also be rewritten as the constraint that every entry of $A\mathcal{P} - b \in \mathbb{R}^{2N}$ is non positive, where A is a $\mathbb{R}^{2N} \times \mathbb{R}^{N^2}$ matrix, and b is a vector in $\mathbb{R}^{2N}$. It follows that $\mathcal{P}$ is a (bounded) convex polytope in $\mathbb{R}^{N^2}$.

If $\pi \in \mathcal{P} \subset \mathbb{R}^{N^2}$ has entries equal to only 0 or 1, then π must be an extremal point of $\mathcal{P}$ by the following argument. If $\pi_\ell = 1$, and π is a nontrivial convex combination of π^1 and π^2 in $\mathcal{P}$, then

$$\pi = \lambda\,\pi^1 + (1 - \lambda)\,\pi^2 \text{ with } \lambda \in (0, 1) \implies 1 = \lambda\,\pi_\ell^1 + (1 - \lambda)\,\pi_\ell^2 \text{ with } \pi_\ell^1,\,\pi_\ell^2 \geq 0 \implies \pi_\ell^1 = \pi_\ell^2 = 1\ .$$

A similar argument can be applied for the entries of π that are 0. It follows that we must have $\pi^1 = \pi = \pi^2$, proving that π is extremal.

It remains thus to prove only that $\mathcal{P}$ has no other extremal points. To achieve this, it suffices to prove that the extremal points of $\mathcal{P}$ are all integer vectors, i.e. vectors all entries of which are integers – once this is established, the Lemma is proved, since the only integer vectors in $\mathcal{P}$ are those with all entries in $\{0, 1\}$.

To prove that the extremal points of $\mathcal{P}$ are all integer vectors, we invoke the Hoffman-Kruskal theorem (see [17], Theorem 7C.1), which states that, given a $L \times K$ matrix $\mathbb{M}$, with all entries in $\{-1, 0, 1\}$, and a vector $b \in \mathbb{R}^L$ with integer entries, the vertices of the polytope defined by $\{x \in \mathbb{R}^K \,;\, (\mathbb{M}x)_\ell \leq b_\ell \text{ for } \ell = 1, \ldots, L\}$ are all integer vectors in $\mathbb{R}^K$ if and only if the matrix $\mathbb{M}$ is totally unimodular, i.e. if and only if every square submatrix of $\mathbb{M}$ has determinant 1, 0 or -1.

We first note that (C.1) can indeed be written in this special form. The equality $\sum_{i,j} \pi_{ij} = M$ can be recast as the two inequalities $\sum_{i,j} \pi_{ij} \leq M$ and $-\sum_{i,j} \pi_{ij} \leq -M$. The full system (C.1) can then be written as $(\mathbb{M}\pi)_\ell \leq b_\ell$ for $\ell = 1, \ldots, L$, where $\mathbb{M}$ is a $(2N + 2 + N^2) \times N^2$ matrix constructed as follows. Its first $2N$ rows correspond to the constraints on the sums over rows and columns; the entries of the next row are all 1, and of the row after that, all -1 – these two rows correspond to the constraint $\sum_{i,j} \pi_{ij} = M$; the final $N^2 \times N^2$ block is diagonal, with all its diagonal entries equal to -1. The first $2N$ entries of b are 1; the next 2 entries are M and $-M$; its final N^2 entries are 0. By the Hoffman-Kruskal theorem it suffices thus to show that $\mathbb{M}$ is totally unimodular.

Because the last N^2 rows, the *bottom rows* of $\mathbb{M}$, have only one non-zero entry, which equals -1, we can disregard them. Indeed, if we take a square submatrix of $\mathbb{M}$ that includes (part of) one of these bottom rows, then the determinant of the submatrix is 0 if only zero entries of the bottom row ended up in the submatrix; if the one -1 entry of the bottom row is an entry in the submatrix, then the determinant is, possibly up to a sign change, the same as if that row and the column of the -1 entry are removed. By this argument, we can remove all the rows of the submatrix partaking of the bottom rows of $\mathbb{M}$.

We thus have to check unimodularity only for $\mathbb{M}'$, the submatrix of $\mathbb{M}$ given by its first $2N + 2$ rows. If any submatrix contains (parts of) both the $(2N+1)$st and the $(2N+2)$nd row, then the determinant is automatically zero, since the second of these two rows equals the first one, multiplied by -1. This reduces the problem to checking that $\mathbb{M}''$, the submatrix of $\mathbb{M}$ given by its first $2N + 1$ rows, is totally unimodular.

We now examine the top $2N$ rows of $\mathbb{M}''$ more closely. A little scrutiny reveals that it is, in fact, the adjacency matrix $\mathbb{G}$ of the complete bipartite graph with N vertices in each part.[1] It is well-known (see e.g. Theorem 8.3 in [29]) that this adjacency matrix is totally unimodular, so any square submatrix of $\mathbb{M}''$ that does not involve the $(2N+1)$st row of $\mathbb{M}''$ is already known to have determinant 0, 1 or -1. We thus have to check only submatrices that involve the last row, i.e. matrices that consist of a $(n-1) \times n$ submatrix of $\mathbb{G}$, with an added nth row with all entries equal to 1. We'll denote such submatrices by $\mathbb{G}'$.

We can then use a simple induction argument on n to finish the proof. The case $n = 2$ is trivial. In proving the induction step for $n = m$, we can assume that each of the top $m - 1$ rows of our $m \times m$ submatrix $\mathbb{G}'$ contains at least two entries equal to 1, since otherwise the determinant of $\mathbb{G}'$ would automatically be 0, 1 or -1 by induction.

The first $m - 1$ rows of $\mathbb{G}'$ correspond to vertices in the bipartite graph, and can thus be partitioned into two sets S_1 and S_2, based on which of the two parts of N vertices in the graph they pertain to. Let us call S the larger of S_1 and S_2; S consists of at least $\lceil \frac{m-1}{2} \rceil$ rows. Let us examine the $(\#S) \times m$ sub-matrix $\mathbb{G}''$ constructed from exactly these rows. We know that each column of $\mathbb{G}''$ has exactly one entry 1, since all the rows of $\mathbb{G}''$ correspond to the same group of vertices in the bipartite graph. Therefore, summing all the rows of $\mathbb{G}''$ gives a vector v of only zeros and ones; since each row in $\mathbb{G}''$ contains at least two entries equal to 1, the sum of all entries in v is at least $2 \left(\lceil \frac{m-1}{2} \rceil \right) \geq m - 1$. The vector v has thus at least $m - 1$ entries equal to 1; the remaining mth entry of this linear combination of the top $m - 1$ rows of $\mathbb{G}'$ is either 1 or 0. In the first case, the determinant of $\mathbb{G}'$ vanishes, since its last row also consists of only ones. In the second case, we can subtract v from the last row of $\mathbb{G}'$ without changing the value of the determinant; the resulting last

---

[1] The adjacency matrix A for a graph $\mathcal{G}$ has as many columns as $\mathcal{G}$ has edges, and as many rows as $\mathcal{G}$ has vertices; if we label the rows and columns of A accordingly, then $A_{ve} = 1$ if the vertex v is an end point of the edge e; otherwise $A_{ve} = 0$. An adjacency matrix thus has exactly two nonzero entries (both equal to 1) in each column. The number of nonzero entries in the row with index v is the degree of v in the graph.

row has all entries but one equal to 0, with a remaining entry equal to 1. The determinant is then given by the minor of this remaining entry, and is thus 0, 1 or -1 by the unimodularity of $\mathbb{G}$. $\square$

APPENDIX D.

In this Appendix we provide a constructive procedure and convergence analysis for approximating the optimal transport cost between *general* separable complete compact metric spaces $(\mathcal{X}, d_\mathcal{X}), (\mathcal{Y}, d_\mathcal{Y})$ each equipped with a probability measure $\mu \in P(\mathcal{X}), \nu \in P(\mathcal{Y})$, where $P(\mathcal{X})$ $(P(\mathcal{Y}))$ denotes the set of probability measures on $\mathcal{X}$ $(\mathcal{Y})$. In the context of the algorithm previously described $\mathcal{X}, \mathcal{Y}$ are the two given surfaces, $d_\mathcal{X}, d_\mathcal{Y}$ the corresponding geodesic distance metric functions, and μ, ν the area measures of the surfaces induced from the metric tensors, respectively. Since R, μ, ν are kept fixed through this discussion, we will denote, for brevity, $c(x, y) = d_{\mu,\nu}^R(x, y)$.

The Kantorovich optimal transportation cost of the measures μ, ν is defined as

$$(\text{D.1}) \qquad T_c(\mu, \nu) = \inf_{\pi \in \Pi(\mu,\nu)} \int_{X \times Y} c(x, y) d\pi(x, y),$$

where $\Pi(\mu, \nu) \subset P(\mathcal{X} \times \mathcal{Y})$ is the set of probability measures on $\mathcal{X} \times \mathcal{Y}$ with marginals μ, ν, that is, $\pi \in \Pi(\mu, \nu) \Rightarrow \pi(A \times \mathcal{Y}) = \mu(A)$ and $\pi(\mathcal{X} \times A') = \nu(A')$, for all Borel $A \subset \mathcal{X}, A' \subset \mathcal{Y}$.

The main goal of this section is to present an *approximation result* for $T_c(\mu, \nu)$ in this general framework. In particular, this result will assure the convergence of our algorithm.

To our knowledge, the only related result talks merely about convergence of the optimal cost (e.g., [32], Theorem 5.20). However, for practical applications it is important to control the *rate* of convergence, and therefore to be able to compute error-bounded approximations.

In the specific case of $c(x, y) = \|x - y\|^2$ there are good approximation techniques that rely on the polar decomposition of Brenier, for example the work of Haker and collaborators [13]. However, as far as we are aware no approximation result is known in the general metric case as required here.

We will show that solving discrete mass-transportation between two sets of discrete measures μ_S, ν_T, based on Voronoi diagrams of two collections of points $S = \{s_i\} \subset \mathcal{X}, T = \{t_j\} \subset \mathcal{Y}$, achieve linear approximation order to the continuous limit mass-transport cost $T_c(\mu, \nu)$:

$$\left| T_c(\mu, \nu) - T_c(\mu_S, \nu_T) \right| \leq \omega_c(2h),$$

where $\omega_c(\alpha)$ is the modulus of continuity of c defined by

$$\omega_c(\alpha) = \sup_{d_\mathcal{X}(x,x')+d_\mathcal{Y}(y,y')<\alpha} |c(x, y) - c(x', y')|,$$

and $h = \max\{\eta(S), \eta(T)\}$, where the *fill distances* $\varphi_\mathcal{X}(S), \varphi_\mathcal{Y}(T)$ are defined as before by

$$(\text{D.2}) \qquad \varphi_\mathcal{X}(S) = \sup \left\{ r \in \mathbb{R} \;\middle|\; \exists x \in \mathcal{X} \text{ s.t. } B_\mathcal{X}(x, r) \cap S = \emptyset \right\},$$

where $B_\mathcal{X}(x, r) = \{q \in \mathcal{X} \mid d_\mathcal{X}(x, q) < r\}$ (and similarly for $\varphi_\mathcal{Y}(T)$).

In particular, for Lipschitz cost function c with Lipschitz constant λ, we have the following bound for the error in the approximation:

$$(\text{D.3}) \qquad \left| T_c(\mu, \nu) - T_c(\mu_S, \nu_T) \right| \leq 2\lambda h.$$

In turn, this result suggests an algorithm for approximating $T_c(\mu, \nu)$: simply spread points $S \subset \mathcal{X}$ and $T \subset \mathcal{Y}$ such that no big empty space is left uncovered, then compute $T_c(\mu_S, \nu_T)$ using linear-programming solver.

D.1. **Voronoi cells and discrete measures.** Let $(\mathcal{X}, d_{\mathcal{X}}, \mu), (\mathcal{Y}, d_{\mathcal{Y}}, \nu)$ be two compact, complete, separable metric spaces with probability measures defined over the Borel sets.

The discrete measures will based on discrete sets of points $S = \{s_i\}_{i=1}^m \subset \mathcal{X}, T = \{t_i\}_{i=1}^n \subset \mathcal{Y}$ and the coarseness of the sets will be measured by means of the so-called *fill-distance* $\varphi_{\mathcal{X}}(S), \varphi_{\mathcal{Y}}(T)$ of the sets S, T. The fill-distance of $S \subset \mathcal{X}$ is defined in eq. (D.2). The sets get "finer" as $\max\{\varphi_{\mathcal{X}}(S), \varphi_{\mathcal{Y}}(T)\} \to 0$.

Definition D.1. For set of points $S = \{s_i\}_{i=1}^m \subset \mathcal{X}$ we define

(D.4)
$$O_i^S = \left\{ x \in \mathcal{X} \ \middle| \ d_{\mathcal{X}}(x, s_i) < \min_{j \neq i} d_{\mathcal{X}}(x, s_j) \right\}.$$

(D.5)
$$C_i^S = \left\{ x \in \mathcal{X} \ \middle| \ d_{\mathcal{X}}(x, s_i) \leq \min_{j \neq i} d_{\mathcal{X}}(x, s_j) \right\}.$$

Then, Voronoi cells for the point set S is any collection of sets $\{V_i\}_{i=1}^m$ satisfying

(1) $\cup_{i=1}^m V_i = \mathcal{X}$.
(2) $V_i \cap V_j = \emptyset$ for $i \neq j$.
(3) $O_i^S \subset V_i \subset C_i^S$, for all $i = 1, .., m$.

We will prove now a simple lemma, for later use, connecting the fill-distance with the geometry of the Voronoi cells.

Lemma D.2. *Let $S = \{s_i\}_{i=1}^m \subset \mathcal{X}$. If $\{V_i\}_{i=1}^m$ is a collection of Voronoi cells corresponding to S, then for all $\varepsilon > 0$,*
$$V_i \subset B(s_i, \varphi_{\mathcal{X}}(S) + \varepsilon), \quad i = 1, .., m$$

Proof. Take $x \notin B(s_i, \varphi_{\mathcal{Y}}(S) + \varepsilon)$.

By the definition of the fill-distance we have that
$$B\left(x, \varphi_{\mathcal{X}}(S) + \frac{\varepsilon}{2}\right) \cap S \neq \emptyset.$$

That is, there exists $s_k \in S$, $k \neq i$, such that
$$d_{\mathcal{X}}(x, s_k) < \varphi_{\mathcal{X}}(S) + \frac{\varepsilon}{2}.$$

However,
$$d_{\mathcal{X}}(x, s_i) \geq \varphi_{\mathcal{X}}(S) + \varepsilon > d_{\mathcal{X}}(x, s_k).$$

Hence from the definition of the Voronoi cells
$$x \in \mathcal{X} \setminus C_i^S \subset \mathcal{X} \setminus V_i,$$

that is $x \notin V_i$. $\square$

Given a set of points and a Voronoi cell collection $S, \{V_i\}_{i=1}^m$ we define a discrete measure μ_S by

(D.6)
$$\mu_S = \sum_{i=1}^m \mu(V_i) \delta_{s_i},$$

where δ_{s_i} is the dirac measure centered at s_i. That is
$$\int_{\mathcal{X}} f \ d\mu_S = \sum_{i=1}^m \mu(V_i) f(s_i).$$

Similarly we define Voronoi cell collection $\{W_j\}_{j=1}^n$ for point set $T \subset \mathcal{Y}$, and corresponding discrete measure ν_T.

Let us prove that the discrete measures μ_S, ν_T converge in the weak sense to μ, ν. By weak convergence $\mu_S \to \mu$ of measure we mean (see for example [23]) that for every continuous bounded function $f : \mathcal{X} \to \mathbb{R}$ there exists

$$\int_{\mathcal{X}} f \, d\mu_S \longrightarrow \int_{\mathcal{X}} f \, d\mu \ , \quad \text{as } \varphi_{\mathcal{X}}(S) \to 0.$$

Theorem D.3. $\lim_{\varphi_{\mathcal{X}}(S) \to 0} \int_{\mathcal{X}} f \, d\mu_S = \int_{\mathcal{X}} f \, d\mu$, for all $f : \mathcal{X} \to \mathbb{R}$ bounded and continuous.

Proof. X is compact therefore f is uniformly continuous. Take arbitrary $\varepsilon > 0$, let $\delta(\varepsilon) > 0$ be such that

$$x, x' \in \mathcal{X} \ , \ d_{\mathcal{X}}(x, x') < \delta(\varepsilon) \ \Rightarrow \ |f(x) - f(x')| < \varepsilon.$$

For S with $\varphi_{\mathcal{X}}(S) < \delta(\varepsilon)$ we have

$$\left| \int_{\mathcal{X}} f \, d\mu - \int_{\mathcal{X}} f \, d\mu_S \right| = \left| \sum_{i=1}^{m} \int_{V_i} (f(x) - f(s_i)) \, d\mu \right| \leq$$

$$\leq \sum_{i=1}^{m} \int_{V_i} |f(x) - f(s_i)| \, d\mu \leq \sum_{i=1}^{m} \varepsilon \mu(V_i) = \varepsilon,$$

where in the second to last transition we used the fact that $V_i \subset B(s_i, \delta(\varepsilon))$ which we know is the case from Lemma D.2. $\square$

Lastly, we will denote by $\Pi(\mu_S, \nu_T) \subset P(S \times T)$ the subset of probability measures on the discrete product space $S \times T$ with marginals μ_S, ν_T.

D.2. **Approximation of optimal cost.** In this subsection we prove the main result of this appendix. Namely, that the optimal transport cost $T_c(\mu_S, \nu_T)$ of the discrete measures μ_S, ν_T is an ε−approximation to the optimal transport cost $T_c(\mu, \nu)$ of μ, ν if the fill-distance $h = \max\{\varphi_{\mathcal{X}}(S), \varphi_{\mathcal{Y}}(T)\} < \frac{1}{2}\delta(\varepsilon)$, where $\delta(\varepsilon)$ is the uniform continuity constant of c.

Actually, we will prove a slightly stronger result: denote the sets

(D.7) $$\mathcal{A} = \left\{ \int_{\mathcal{X} \times \mathcal{Y}} c \, d\pi \ \Big| \ \pi \in \Pi(\mu, \nu) \right\} \ , \quad \mathcal{B} = \left\{ \int_{\mathcal{X} \times \mathcal{Y}} c \, d\pi \ \Big| \ \pi \in \Pi(\mu_S, \nu_T) \right\}.$$

We will show that the Hausdorff distance $d_H(\mathcal{A}, \mathcal{B}) \to 0$ as $h \to 0$. Where by Hausdorff distance of two sets $\mathcal{A}, \mathcal{B} \subset \mathbb{R}$ we mean

$$d_H(\mathcal{A}, \mathcal{B}) = \inf \left\{ r \ \Big| \ \mathcal{B} \subset U(\mathcal{A}, r) \ , \ \mathcal{A} \subset U(\mathcal{B}, r) \right\},$$

where $U(\mathcal{A}, r) = \cup_{a \in \mathcal{A}} B(a, r)$ and $B(a, r)$ is the open ball of radius r centered at a. Moreover, we will provide a linear (in h) bound controlling the convergence rate, for Lipschitz cost function c.

Theorem D.4. If $c : \mathcal{X} \times \mathcal{Y} \longrightarrow \mathbb{R}_+$ is a continuous function, $\mathcal{X}, \mathcal{Y}$ compact complete separable metric spaces. Let $h = \max\{\varphi_{\mathcal{X}}(S), \varphi_{\mathcal{Y}}(T)\}$ then,

$$d_H(\mathcal{A}, \mathcal{B}) \leq \omega_c(2h),$$

where $\mathcal{A}, \mathcal{B}$ are defined in (D.7). In particular,

$$\left| T_c(\mu, \nu) - T_c(\mu_S, \nu_T) \right| \leq \omega_c(2h).$$

Proof. Take arbitrary $S = \{s_i\}_{i=1}^{m} \subset \mathcal{X}$, $T = \{t_i\}_{j=1}^{n} \subset \mathcal{Y}$. Choose collections of Voronoi cells $\{V_i\}_{i=1}^{m}$, $\{W_j\}_{j=1}^{n}$ for S, T respectively.

Now take $\pi \in \Pi(\mu, \nu)$.

Set

$$\Lambda = \sum_{i=1}^{m} \sum_{j=1}^{n} \pi(V_i \times W_j) \delta_{s_i, t_j}.$$

Then,

$$\Lambda(A \times \mathcal{Y}) = \sum_{i=1}^{m} \sum_{j=1}^{n} \pi(V_i \times W_j) \delta_{s_i}(A) =$$

$$\sum_{i=1}^{m} \pi(V_i \times \mathcal{Y}) \delta_{s_i}(A) = \sum_{i=1}^{m} \mu(V_i) \delta_{s_i}(A) = \mu_S(A).$$

Similarly

$$\Lambda(\mathcal{X} \times A') = \nu_T(A').$$

Therefore

$$\Lambda \in \Pi(\mu_S, \nu_T).$$

Moreover,

$$\left| \int_{\mathcal{X} \times \mathcal{Y}} c \, d\pi - \int_{\mathcal{X} \times \mathcal{Y}} c \, d\Lambda \right| = \left| \sum_{i=1}^{m} \sum_{j=1}^{n} \int_{V_i \times W_j} \left[c(x,y) - c(s_i,t_j) \right] \, d\pi \right| \leq$$

$$\sum_{i=1}^{m} \sum_{j=1}^{n} \int_{V_i \times W_j} \left| c(x,y) - c(s_i,t_j) \right| \, d\pi.$$

For $(x,y) \in V_i \times W_j$, we have $d_{\mathcal{X}}(x,s_i) \leq \varphi_{\mathcal{X}}(S)$, $d_{\mathcal{Y}}(y,t_j) \leq \varphi_{\mathcal{Y}}(T)$ which implies

(D.8) $$|c(x,y) - c(s_i,t_j)| \leq \omega_c \left(\varphi_{\mathcal{X}}(S) + \varphi_{\mathcal{Y}}(T) \right) \leq \omega_c(2h).$$

So we have

$$\left| \int_{\mathcal{X} \times \mathcal{Y}} c \, d\pi - \int_{\mathcal{X} \times \mathcal{Y}} c \, d\Lambda \right| \leq \omega_c(2h) \sum_{i=1}^{n} \sum_{j=1}^{n} \pi(V_i \times W_j) = \omega_c(2h),$$

since $\{V_i \times W_j\}_{i,j=1,1}^{n,m}$ form a partition of $X \times Y$. So we proved that

$$\mathcal{A} \subset \overline{U(\mathcal{B}, \omega_c(2h))}.$$

We now prove the other direction. Take $\Lambda \in \Pi(\mu_S, \nu_T)$.

Denote by $\mu \times \nu \big|_{V_i \times W_j}$ the product measure $\mu \times \nu$ restricted to $V_i \times W_j \subset \mathcal{X} \times \mathcal{Y}$.

That is,

$$\mu \times \nu \big|_{V_i \times W_j} (A \times A') = \mu \times \nu \left((V_i \times W_j) \cap (A \times A') \right) = \mu(V_i \cap A) \nu(W_j \cap A').$$

Now pick

$$\pi = \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{\Lambda(V_i \times W_j)}{\mu(V_i) \nu(W_j)} \mu \times \nu \big|_{V_i \times W_j}.$$

Then,

$$\pi(A \times \mathcal{Y}) = \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{\Lambda(V_i \times W_j)}{\mu(V_i) \nu(W_j)} \mu(V_i \cap A) \nu(W_j \cap \mathcal{Y}) =$$

$$\sum_{i=1}^{m} \sum_{j=1}^{n} \frac{\Lambda(V_i \times W_j)}{\mu(V_i)} \mu(V_i \cap A) =$$

$$\sum_{i=1}^{m} \frac{\mu(V_i \cap A)}{\mu(V_i)} \left[\sum_{j=1}^{n} \Lambda(V_i \times W_j) \right] =$$

$$\sum_{i=1}^{m} \frac{\mu(V_i \cap A)}{\mu(V_i)} \Lambda(V_i \times \mathcal{Y}) =$$

$$\sum_{i=1}^{m} \frac{\mu(V_i \cap A)}{\mu(V_i)} \mu(V_i) =$$

$$\sum_{i=1}^{m} \mu(V_i \cap A) = \mu(\mathcal{X} \cap A) = \mu(A).$$

Similarly

$$\pi(\mathcal{X} \times A') = \nu(A').$$

Therefore,

$$\pi \in \Pi(\mu, \nu).$$

Now,

$$\left| \int_{\mathcal{X} \times \mathcal{Y}} c \, d\pi - \int_{\mathcal{X} \times \mathcal{Y}} c \, d\Lambda \right| =$$

$$\left| \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{\Lambda(V_i \times W_j)}{\mu(V_i)\nu(W_j)} \int_{V_i \times W_j} c \, d(\mu \times \nu) - \sum_{i=1}^{m} \sum_{j=1}^{n} c(s_i, t_j) \Lambda(V_i \times W_j) \right| =$$

$$\left| \sum_{i=1}^{m} \sum_{j=1}^{n} \left[\frac{\Lambda(V_i \times W_j)}{\mu(V_i)\nu(W_j)} \int_{V_i \times W_j} c \, d(\mu \times \nu) - \frac{\Lambda(V_i \times W_j)}{\mu(V_i)\nu(W_j)} \int_{V_i \times W_j} c(s_i, t_j) \, d(\mu \times \nu) \right] \right| \le$$

$$\sum_{i=1}^{m} \sum_{j=1}^{n} \frac{\Lambda(V_i \times W_j)}{\mu(V_i)\nu(W_j)} \int_{V_i \times W_j} |c(x, y) - c(s_i, t_j)| \, d(\mu \times \nu).$$

As before, for $(x, y) \in V_i \times W_j$, we have $d_{\mathcal{X}}(x, s_i) \le \varphi_{\mathcal{Y}}(S)$, $d_{\mathcal{Y}}(y, t_j) \le \varphi_{\mathcal{Y}}(T)$ therefore

$$|c(x, y) - c(s_i, t_j)| \le \omega_c(\varphi_{\mathcal{X}}(S) + \varphi_{\mathcal{Y}}(T)) \le \omega_c(2h).$$

And therefore

$$\left| \int_{\mathcal{X} \times \mathcal{Y}} c \, d\pi - \int_{\mathcal{X} \times \mathcal{Y}} c \, d\Lambda \right| \le \omega_c(2h) \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{\Lambda(V_i \times W_j)}{\mu(V_i)\nu(W_j)} \mu(V_i)\nu(W_j) =$$

$$\omega_c(2h) \sum_{i=1}^{m} \sum_{j=1}^{n} \Lambda(V_i \times W_j) = \omega_c(2h).$$

So we proved

$$\mathcal{B} \subset \overline{U(\mathcal{A}, \omega_c(2h))},$$

and

$$d_H(\mathcal{A}, \mathcal{B}) \le \omega_c(2h).$$

In particular this means that

$$\left| T_c(\mu, \nu) - T_c(\mu_S, \nu_T) \right| = \left| \inf(\mathcal{A}) - \inf(\mathcal{B}) \right| \le \omega_c(2h).$$

$$\square$$

We will call $c : \mathcal{X} \times \mathcal{Y} \longrightarrow \mathbb{R}_+$ Lipschitz continuous with a constant λ if

$$|c(x, y) - c(x', y')| \le \lambda \left(d_{\mathcal{X}}(x, x') + d_{\mathcal{Y}}(y, y') \right), \ \forall x, x' \in \mathcal{X}, \ y, y' \in \mathcal{Y}.$$

Theorem 2.3 *Suppose $c : \mathcal{X} \times \mathcal{Y} \longrightarrow \mathbb{R}_+$ is a continuous function, with $\mathcal{X}, \mathcal{Y}$ compact complete separable metric spaces, S and T are sample sets in $\mathcal{X}, \mathcal{Y}$ (resp.), μ, ν are probability measures on $\mathcal{X}, \mathcal{Y}$.*
(A) if c is uniformly continuous then

$$T_c(\mu_S, \nu_T) \to T_c(\mu, \nu), \quad as \ h \to 0,$$

(B) if c is Lipschitz continuous with a constant λ, then

$$|T_c(\mu, \nu) - T_c(\mu_S, \nu_T)| < 2\lambda h,$$

where, $h = \max\{\varphi_{\mathcal{X}}(S), \varphi_{\mathcal{Y}}(T)\}$, and μ_S, ν_T are as defined in (D.6).

Proof. The result follow from the fact that for uniformly continuous c, $\omega_c(t) \to 0$ as $t \to 0$, and that for Lipschitz c with constant λ, $\omega_c(t) \leq \lambda t$. $\qquad \square$

Therefore, a simple generalization to the non-compact case, that is when $\mathcal{X}, \mathcal{Y}$ are complete separable metric spaces, can be achieved by requiring that $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}_+$ is uniformly continuous on $\mathcal{X} \times \mathcal{Y}$.

Corollary D.5. *For $\mathcal{X}, \mathcal{Y}$ complete, separable metric spaces, and $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}_+$ uniformly continuous*

$$T_c(\mu_S, \nu_T) \longrightarrow T_c(\mu, \nu).$$

Moreover, for Lipschitz c with Lipschitz constant λ

$$|T_c(\mu, \nu) - T_c(\mu_S, \nu_T)| < 2\lambda h,$$

where, as before, $h = \max\{\varphi_{\mathcal{X}}(S), \varphi_{\mathcal{Y}}(T)\}$.

Note that our argument used only one specific property of Voronoi cells, namely that each is contained in an h-size closed ball. Many other ways to partition $\mathcal{X}$ and $\mathcal{Y}$ can be considered. The next Lemma uses a property of the Voronoi cells to show that the proposed discretization with Voronoi cells is in some sense optimal.

Lemma D.6. *Let $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}_+$ be Lipschitz with constant λ, and let $S = \{s_i\}_{i=1}^m \subset \mathcal{X}$, $T = \{t_j\}_{j=1}^n \subset \mathcal{Y}$ be given point sets. Then, among all the choices of subdividing $\mathcal{X}$ and $\mathcal{Y}$, $\mathcal{X} = \cup_{i=1}^m Q_i$, and $\mathcal{Y} = \cup_{j=1}^n R_j$, the Voronoi cells $Q_i = V_i, R_j = W_j$ minimize a bound on the error term:*

$$\left| \int_{\mathcal{X} \times \mathcal{Y}} c \, d\pi - \int_{\mathcal{X} \times \mathcal{Y}} c \, d\Lambda \right|.$$

Proof.

$$\left| \int_{\mathcal{X} \times \mathcal{Y}} c \, d\pi - \int_{\mathcal{X} \times \mathcal{Y}} c \, d\Lambda \right| \leq \sum_{i=1}^m \sum_{j=1}^n \int_{Q_i \times R_j} |c(x,y) - c(s_i, t_j)| \, d\pi \leq$$

$$\sum_{i=1}^m \sum_{j=1}^n \int_{Q_i \times R_j} \lambda \left[d_{\mathcal{X}}(x, s_i) + d_{\mathcal{Y}}(y, t_j) \right] d\pi =$$

$$\lambda \sum_{i=1}^m \int_{Q_i \times \mathcal{Y}} d_{\mathcal{X}}(x, s_i) d\pi + \lambda \sum_{j=1}^n \int_{\mathcal{X} \times R_j} d_{\mathcal{X}}(y, t_j) d\pi =$$

$$\lambda \sum_{i=1}^m \int_{Q_i} d_{\mathcal{X}}(x, s_i) d\nu + \lambda \sum_{j=1}^n \int_{R_j} d_{\mathcal{X}}(y, t_j) d\mu,$$

and it is not hard to see that the choices of R_i, Q_j that minimizes this last error bound are the Voronoi cells $R_i = V_i$, and $Q_j = W_j$, where $\{V_i\}, \{W_j\}$ are the Voronoi cells corresponding to S, T, respectively. $\qquad \square$