ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Advances in Mathematics 227 (2011) 1047–1077

ADVANCES IN Mathematics

www.elsevier.com/locate/aim

Conformal Wasserstein distances: Comparing surfaces in polynomial time

Y. Lipman *, I. Daubechies

Princeton University, United States

Received 20 January 2011; accepted 25 January 2011

Available online 14 April 2011

Communicated by Gang Tian

Abstract

We present a constructive approach to surface comparison realizable by a polynomial-time algorithm. We determine the "similarity" of two given surfaces by solving a mass-transportation problem between their conformal densities. This mass transportation problem differs from the standard case in that we require the solution to be invariant under global Möbius transformations. We present in detail the case where the surfaces to compare are disk-like; we also sketch how the approach can be generalized to other types of surfaces.

Keywords: Numerical analysis; Differential geometry; Mass transportation; Surface comparison

1. Introduction

Alignment and comparison of surfaces (2-manifolds) play a central role in a wide range of scientific disciplines; they often constitute a crucial step in a variety of problems in medicine and biology.

Mathematically, the algorithmic problem of surface alignment amounts to defining a metric function $\mathbf{d}(\cdot\,,\cdot)$ in the space of Riemannian 2-manifolds with the following two properties:

1) for any two surfaces $\mathcal{M}$ and $\mathcal{N}$, $\mathbf{d}(\mathcal{M},\mathcal{N})=0$ implies that $\mathcal{M}$ and $\mathcal{N}$ are isometric, and

* Corresponding author.
E-mail address: ylipman@princeton.edu (Y. Lipman).

0001-8708/$ – see front matter
doi:10.1016/j.aim.2011.01.020

2) given a reasonably large number of reasonably well-distributed sample points on both surfaces, an accurate approximation to the distance $\mathbf{d}(\mathcal{M}, \mathcal{N})$ can be calculated in a time that grows only polynomially in the sample set size.

This second requirement is crucial to ensure that the algorithm can be used effectively in applications.

A prominent mathematical approach to define distances between surfaces that has been proposed for practical applications [10,6] is the *Gromov–Hausdorff (GH) distance*; it considers the surfaces as special cases of *metric spaces*. To determine the GH distance between the metric spaces X and Y, one examines all the isometric embeddings of X and Y into (other) metric spaces; although this distance possesses many attractive mathematical properties, it is inherently hard computationally. For instance, computing the L_p version of the GH distance between two surfaces is equivalent to a non-convex quadratic programming problem, generally over the integers [3]. This problem is equivalent to integer quadratic assignment, and is thus NP-hard [2]. In [3], Memoli generalizes the GH distance of [10] by introducing a quadratic mass transportation scheme to be applied to metric spaces that are also equipped with a measure (mm spaces); the computation of this *Gromov–Wasserstein (GW) distance for mm spaces* is somewhat easier and more stable to implement than the original GH distance. The computation of the GW distance between two surfaces described in [3] utilizes a (continuous rather than integer) quadratic programming method; the functional to be minimized is generally not convex and optimization methods are likely to find local minima rather than the global minima that realizes the surfaces' distance.

In this paper we propose a new surface alignment procedure, introducing the *conformal Wasserstein distance*. Our construction consists in "geometrically" aligning the surfaces, based on uniformization theory and optimal mass transportation. The uniformization theory serves as a "dimensionality reduction" tool, representing e.g. a disk-type surface by its conformal factor on the unit disk: the corresponding automorphism group (the disk-preserving Möbius group) has only three degrees of freedom and is therefore searchable in polynomial time. Next, the Kantorovich mass-transportation [5] is used to construct a linear functional the minimizer of which furnishes a metric; as is well known [12,14], this can solved by a linear program, and can thus be computed/approximated in polynomial time as well.

As far as we know, prior to our work, no polynomial time algorithm was known to compute, either exactly or up to a good approximation, the GH distance or any other proposed intrinsic geometric distances between surfaces. Although [10] uses a mass transportation as well (albeit quadratic mass transportation), our approach is nevertheless different. We solve the "standard" (and thus *linear*) Kantorovich mass transportation problem, which is convex (even linear) and solvable via a linear programming method.

There exist earlier papers on aligning or comparing surfaces that use uniformization. In particular, the papers by Zeng et al. [15,16] which build upon the work of Gu and Yau [4], also use uniformization for surface alignment (albeit without defining a distance between surfaces). However, 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. We use only intrinsic geometric information: we make use of the surfaces' metric (inherited from its embedding in $\mathbb{R}^3$) and the induced conformal structure to define deviation from (local) isometry.

Optimal mass transportation has also been used before in aligning or comparing images. Following the seminal work by Rubner et al. [12], it is used extensively in the engineering literature to define interesting metric distances for images, interpreted as probability densities; in this context the metric is often called the "Earth Mover's Distance".

Our paper is organized as follows: in Section 2 we briefly recall some facts about uniformization and optimal mass transportation that we shall use, at the same time introducing our notation. Section 3 contains the main results of this paper, constructing the conformal Wasserstein distance metric between disk-type surfaces, in several steps; we also indicate how the approach can be generalized to other surfaces. Section 4 briefly describes the discrete case.

2. Background and notations

As described in the introduction, our framework makes use of two mathematical theories: uniformization theory, to represent the surfaces as measures defined on a canonical domain, and optimal mass transportation, to align the measures. In this section we recall some of their basic properties, and we introduce our notations.

2.1. Uniformization

By the celebrated uniformization theory for Riemann surfaces (see for example [13,7]), any simply-connected Riemann surface is conformally equivalent to one of three canonical domains: the sphere, the complex plane, or the unit disk. Since every 2-manifold surface $\mathcal{M}$ equipped with a smooth Riemannian metric g has an induced conformal structure and is thus a Riemann surface, uniformization applies to such surfaces. Therefore, every simply-connected surface with a Riemannian metric can be mapped conformally to one of the three canonical domains listed above. In this paper, we discuss 2D surfaces, equipped with a Riemannian metric tensor g (possibly inherited from the standard Euclidean 3D metric if the surface is embedded in $\mathbb{R}^3$) that have a finite total volume (i.e. area, since we are dealing with surfaces). For convenience, we shall normalize the metric so that the surface area equals 1. We shall discuss in detail the case where the surfaces $\mathcal{M}$ are topologically equivalent to disks. (We shall address in side remarks how the approach can be extended to the other cases.) For each such $\mathcal{M}$ there exists a conformal map $\phi : \mathcal{M} \to \mathcal{D}$, where $\mathcal{D} = \{z;\ |z| < 1\}$ is the open unit disk. (we assume that $\mathcal{M}$ does not include its boundary, if it has one). The map ϕ pushes g to a metric on $\mathcal{D}$; denoting the coordinates in $\mathcal{D}$ by $z = x^1 + \mathbf{i}x^2$, we can write this metric as

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

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

$$\mathrm{vol}_{\mathcal{M}}(A) = \int_{\phi(A)} \widetilde{\mu}(z)\, dx^1 \wedge dx^2. \tag{2.1}$$

It will be convenient to use the hyperbolic metric $(1 - |z|^2)^{-2}\delta_{ij}\, dx^i \otimes dx^j$ as the reference metric on the unit disk, rather than the standard Euclidean $\delta_{ij}\, dx^i \otimes dx^j$; note that the two are conformally equivalent (with conformal factor $(1-|z|^2)^{-2}$). Instead of the density $\widetilde{\mu}(z)$, we shall therefore use the *hyperbolic density function*

$$\mu^H(z) := \left(1-|z|^2\right)^2 \widetilde{\mu}(z), \tag{2.2}$$

where the superscript H stands for hyperbolic. We shall often drop this superscript: unless otherwise stated $\mu = \mu^H$, and $\nu = \nu^H$. This density function μ satisfies

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

where $d_{\mathrm{vol}_H}(z) = (1-|z|^2)^{-2}\, dx^1 \wedge dx^2$. In what follows we shall use the symbol μ both for the function μ^H and as a shorthand for the absolutely continuous measure $\mathrm{vol}_{\mathcal{M}}$, and by extension for the surface $\mathcal{M}$ itself.

The conformal mappings of $\mathcal{D}$ to itself are the disk-preserving Möbius transformations $m \in M_{\mathcal{D}}$, a family with three real parameters, defined by

$$m(z) = e^{\mathbf{i}\theta} \frac{z-a}{1-\bar{a}z}, \quad a \in \mathcal{D},\ \theta \in [0, 2\pi). \tag{2.3}$$

Since these Möbius transformations satisfy

$$\left(1-\left|m(z)\right|^2\right)^{-2} \left|m'(z)\right|^2 = \left(1-|z|^2\right)^{-2}, \tag{2.4}$$

where m' stands for the derivative of m, the pull-back of μ under a mapping $m \in M_{\mathcal{D}}$ takes on a particularly simple expression. Setting $w = m(z)$, with $w = y^1 + \mathbf{i}y^2$, and $\widetilde{g}(w) = \widetilde{\mu}(w)\delta_{ij}\, dy^i \otimes dy^j = \mu(w)(1-|w|^2)^{-2}\delta_{ij}\, dy^i \otimes dy^j$, the definition

$$\left(m^*\widetilde{g}\right)(z)_{kl}\, dx^k \otimes dx^\ell := \mu(w)\left(1-|w|^2\right)^{-2} \delta_{ij}\, dy^i \otimes dy^j$$

implies

$$\begin{aligned}
\left(m^*\widetilde{g}\right)_{k\ell}(z)\, dx^k \otimes dx^\ell &= \mu\left(m(z)\right)\left(1-\left|m(z)\right|^2\right)^{-2} \delta_{ij} \frac{\partial y^i}{\partial x^k} \frac{\partial y^j}{\partial x^\ell}\, dx^k \otimes dx^\ell \\
&= \mu\left(m(z)\right)\left(1-\left|m(z)\right|^2\right)^{-2} \left|m'(z)\right|^2 \delta_{k\ell}\, dx^k \otimes dx^\ell \\
&= \mu\left(m(z)\right)\left(1-|z|^2\right)^{-2} \delta_{k\ell}\, dx^k \otimes dx^\ell.
\end{aligned}$$

In other words, $(m^*\widetilde{g})(z)_{kl}\, dx^k \otimes dx^\ell$ takes on the simple form

$$m^*\mu(z)\left(1-|z|^2\right)^{-2} \delta_{kl}\, dx^k \otimes dx^\ell,$$

with

$$m^*\mu(z) = \mu\left(m(z)\right). \tag{2.5}$$

Likewise, the push-forward, under a disk Möbius transform $m(z) = w$, of the (diagonal) Riemannian metric defined by the density function $\mu = \mu^H$, is again a diagonal metric, with (hyperbolic) density function $m_*\mu(w) = (m_*\mu)^H(w)$ given by

$$m_*\mu(w) = \mu\big(m^{-1}(w)\big). \tag{2.6}$$

It follows that checking whether or not two surfaces $\mathcal{M}$ and $\mathcal{N}$ are isometric, or searching for (near-)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 $\nu(m(z))$ and $\mu(z)$ coincide (or "nearly" coincide, in a sense to be made precise). This was exploited in [9] to construct fast algorithms to find corresponding points between two given surfaces. In the next section we provide a precise formalization of this idea using the notion of *optimal mass transportation*, described in the following subsection.

2.2. Optimal mass transportation

Optimal mass transportation was introduced by G. Monge [11], and L. Kantorovich [5]. It concerns the transformation of one mass distribution into another while minimizing a cost function that can be viewed as the amount of work required for the task. In the Kantorovich formulation, to which we shall stick in this paper, one considers two measure spaces X, Y (each equipped with a σ-algebra), a probability measure on each, $\mu \in P(X)$, $\nu \in P(Y)$ (where $P(X)$, $P(Y)$ are the respective spaces of all probability measures on X and Y), and the space $\Pi(\mu,\nu)$ of probability measures π on $X \times Y$ with marginals μ and ν (resp.), that is, for $A \subset X$, $B \subset Y$, $\pi(A \times Y) = \mu(A)$ and $\pi(X \times B) = \nu(B)$. The *optimal* mass transportation is the element of $\Pi(\mu,\nu)$ that minimizes $\int_{X\times Y} d(x,y)\,d\pi(x,y)$, where $d(x,y)$ is a cost function. (In general, one should consider an infimum rather than a minimum; in our case, X and Y are compact, $d(\cdot,\cdot)$ is continuous, and the infimum is achieved.) The corresponding minimum,

$$T_d(\mu,\nu) = \inf_{\pi\in\Pi(\mu,\nu)} \int\limits_{X\times Y} d(x,y)\,d\pi(x,y), \tag{2.7}$$

is the optimal mass transportation distance between μ and ν, with respect to the cost function $d(x,y)$.

Intuitively, one can interpret this as follows: imagine being confronted with a pile of sand on the one hand (corresponding to μ), and a hole in the ground on the other hand ($-\nu$), and assume that the volume of the sand pile equals exactly the volume of the hole (suitably normalized, μ,ν are probability measures). You wish to fill the hole with the sand from the pile ($\pi \in \Pi(\mu,\nu)$), in a way that minimizes the amount of work (represented by $\int d(x,y)\,d\pi(x,y)$, where $d(\cdot,\cdot)$ can be thought of as a distance function).

In what follows, we shall apply this framework to the density functions μ and ν on the hyperbolic disk $\mathcal{D}$ obtained by conformal mappings from two surfaces $\mathcal{M},\mathcal{N}$, as described in the previous subsection.

The Kantorovich transportation framework cannot be applied directly to the densities μ,ν. Indeed, the density μ, characterizing the Riemannian metric on $\mathcal{D}$ obtained by pushing forward the metric on $\mathcal{M}$ via the uniformizing map $\phi : \mathcal{M} \to \mathcal{D}$, is not uniquely defined: another uniformizing map $\phi' : \mathcal{M} \to \mathcal{D}$ may well produce a different μ'. Because the two representations are necessarily isometric ($\phi^{-1} \circ \phi'$ maps $\mathcal{M}$ isometrically to itself), we must have $\mu'(m(z)) = \mu(z)$ for some $m \in M_{\mathcal{D}}$. (In fact, $m = \phi' \circ \phi^{-1}$.) In a sense, the representation of (disk-type) surfaces $\mathcal{M}$ as measures over $\mathcal{D}$ should be considered "modulo" the disk Möbius transformations.

We thus need to address how to adapt the optimal transportation framework to factor out this Möbius transformation ambiguity. The next section starts by showing how this can be done.

3. The conformal Wasserstein framework: optimal mass transportation for surfaces

We want to measure distances between surfaces by using the Kantorovich transportation framework to measure the transportation between the metric densities on $\mathcal{D}$ obtained by uniformization applied to the surfaces. The main obstacle is that these metric densities are not uniquely defined; they are defined up to a Möbius transformation. In particular, if two densities μ and ν are related by $\nu = m_*\mu$ (i.e. $\mu(z) = \nu(m(z))$), where $m \in M_{\mathcal{D}}$, then we want our putative distance between μ and ν to be zero, since they describe isometric surfaces, and could have been obtained by different uniformization maps of the same surface. We thus want a distance metric between *orbits* of the group $M_{\mathcal{D}}$ acting on the conformal factors rather than a metric distance between the conformal factors themselves. If we choose a metric distance d on $\mathcal{D}$ that is invariant under Möbius transformations, i.e. that it is a multiple of the hyperbolic distance on the disk, then a natural definition is as follows

$$\mathfrak{D}(\mu, \nu) = \inf_{m \in M_{\mathcal{D}}} \left(\inf_{\pi \in \Pi(m_*\mu, \nu)} \int_{\mathcal{D} \times \mathcal{D}} d(z, w)\, d\pi(z, w) \right). \tag{3.1}$$

As shown in Appendix A, $\mathfrak{D}(\mu, \nu)$ is indeed a distance between disk-type surfaces; its computation can moreover be implemented in running times that grow only polynomially in the number N of sample points used in the discretization of the surface (necessary to proceed to numerical computation). The optimization over m in the definition of $\mathfrak{D}(\mu, \nu)$ always achieves its minimum in some m (depending on μ and ν of course); denoting this special minimizing $m \in M_{\mathcal{D}}$ by $m_{\mu,\nu}$, we can rewrite $\mathfrak{D}(\mu, \nu)$ as the result of a single minimization (for details, see Appendix A):

$$\begin{aligned} \mathfrak{D}(\mu, \nu) &= \inf_{\pi \in \Pi([m_{\mu,\nu}]_*\mu, \nu)} \int_{\mathcal{D} \times \mathcal{D}} d(z, w)\, d\pi(z, w) \\ &= \inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathcal{D} \times \mathcal{D}} d\big(m_{\mu,\nu}(z), w\big)\, d\pi(z, w). \end{aligned} \tag{3.2}$$

This is, however, purely formal; since the determination of $m_{\mu,\nu}$ involves the original double minimization of (3.1), a numerical implementation does require solving a mass-transportation functional for many $m \in M_{\mathcal{D}}$. In practice, this means that, despite its polynomial running time complexity, the numerical computation of $\mathfrak{D}(\mu, \nu)$ is too heavy for many applications, in which all pairwise distances must be computed for a collection that may contain hundreds of surfaces [1]. This seems to lead to an impasse, since there exists no other distance metric on $\mathcal{D}$ that is conformally invariant, so that the natural "quotienting operation" over the group $M_{\mathcal{D}}$ can produce no other metric than $\mathfrak{D}(\cdot\,, \cdot)$.

However, $\mathcal{D}(\mu, \nu)$, rewritten as in (3.2), suggests another way in which we can define an appropriate distance metric between orbits of the group $M_{\mathcal{D}}$ acting on the conformal factors. Note that (3.2) has exactly the same form as for a standard Kantorovich mass transportation scheme, except for the (crucial) difference that *the cost function depends on* μ *and* ν. By retaining the idea of (Kantorovich) mass transportation, but allowing the use of cost functions $d(\cdot, \cdot)$ in the integrand that depend on μ and ν (without picking them necessarily of the form $d(m_{\mu,\nu}(z), w)$), we can construct other distance metrics D on the conformal factors that are invariant under action of $M_{\mathcal{D}}$, i.e. for which $D(\mu, \nu) = D(m_*\mu, \nu)$ for all $m \in M_{\mathcal{D}}$. In addition, we can pick cost functions of this type ensuring that the distance between (or dissimilarity of) μ and ν exhibits some robustness with respect to deviations from global isometry. More precisely, we want the distance to be small for surfaces that are not isometric but nevertheless very close to isometric on *parts* of the surfaces; this can be achieved by picking a cost function $d^R_{\mu,\nu}(z, w)$ that depends on a comparison of the behavior of μ and ν on *neighborhoods* of z and w, mapped by m ranging over $M_{\mathcal{D}}$. This cost function, once incorporated in the Kantorovich mass transportation framework, will lead to a metric between disk-type surfaces (some generic conditions aside) based on solving a *single* mass transportation problem. The next subsection shows precisely how this is done. As is the case throughout the paper, we first give the full details of the construction for disk-like surfaces, and then indicate later how to generalize this to e.g. sphere-like surfaces. It is worthwhile to note that the "quotient approach" sketched above would not even have been applicable in a straightforward way to sphere-like surfaces, since they do not possess a metric invariant under all their Möbius transformations. As we shall explain at the end of this section, the same obstruction will not exist for the construction introduced in the next subsection.

3.1. Construction of $d^R_{\mu,\nu}(z, w)$

We construct $d^R_{\mu,\nu}(z, w)$ so that it 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. We will need to define two ingredients for this: the neighborhoods we will use, and how we shall characterize the (dis)similarity of two neighborhoods, equipped with different metrics.

We start with the neighborhoods.

For a fixed radius $R > 0$, we define $\Omega_{z_0,R}$ to be the hyperbolic geodesic disk of radius R centered at z_0. 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: $\Omega_{0,R} = \{z;\ |z| \leqslant r_R\}$, where $r_R = \tanh(R)$. The hyperbolic disks around other centers are images of these central disks under Möbius transformations (= hyperbolic isometries): setting $m(z) = (z - z_0)(1 - z\overline{z_0})^{-1}$, we have

$$\Omega_{z_0,R} = m^{-1}(\Omega_{0,R}). \tag{3.3}$$

If m', m'' are two maps in $M_{\mathcal{D}}$ that both map z_0 to 0, then $m'' \circ (m')^{-1}$ simply rotates $\Omega_{0,R}$ around its center, over some angle θ determined by m' and m''. From this observation one easily checks that (3.3) holds for *any* $m \in M_{\mathcal{D}}$ that maps z_0 to 0. In fact, we have the following more general

Lemma 3.1. *For arbitrary $z, w \in \mathcal{D}$ and any $R > 0$, every disk preserving Möbius transformation $m \in M_{\mathcal{D}}$ that maps z to w (i.e. $w = m(z)$) also maps $\Omega_{z,R}$ to $\Omega_{w,R}$.*

Next we define how to quantify the (dis)similarity of the pairs $(\Omega_{z_0,R},\mu)$ and $(\Omega_{w_0,R},\nu)$. Since (global) isometries are given by the elements of the disk-preserving Möbius group $M_{\mathcal{D}}$, we will test the extent to which the two patches are isometric by comparing $(\Omega_{w_0,R},\nu)$ with all the images of $(\Omega_{z_0,R},\mu)$ under Möbius transformations in $M_{\mathcal{D}}$ that take z_0 to w_0.

To carry out this comparison, we need a norm. Any metric $g_{ij}(z)\,dx^i \otimes dx^j$ induces an inner product on the space of 2-covariant tensors, as follows: if $\mathbf{a}(z) = a_{ij}(z)\,dx^i \otimes dx^j$ and $\mathbf{b}(z) = b_{ij}(z)\,dx^i \otimes dx^j$ are two 2-covariant tensors in our parameter space $\mathcal{D}$, then their inner product is defined by

$$\langle \mathbf{a}(z), \mathbf{b}(z)\rangle = a_{ij}(z) b_{k\ell}(z) g^{ik}(z) g^{j\ell}(z); \tag{3.4}$$

as always, this inner product defines a norm, $\|\mathbf{a}\|_z^2 = a_{ij}(z)a_{k\ell}(z)g^{ik}(z)g^{j\ell}(z)$. Let us apply this to the computation of the norm of the difference between the local metric on one surface, $g_{ij}(z) = \mu(z)(1-|z|^2)^{-2}\delta_{ij}$, and $h_{ij}(w) = \nu(w)(1-|w|^2)^{-2}\delta_{ij}$, the pull-back metric from the other surface by a Möbius transformation m. Using (3.4), (2.5), and writing $\boldsymbol{\delta}$, $\boldsymbol{g}$, $\boldsymbol{h}$, for the tensors with entries δ_{ij}, g_{ij}, and h_{ij}, respectively, we have:

$$\begin{aligned}\|\boldsymbol{g} - m^*\boldsymbol{h}\|_z^2 &= \left\|\mu(z)\left(1-|z|^2\right)^{-2}\boldsymbol{\delta} - \nu\left(m(z)\right)\left(1-|z|^2\right)^{-2}\boldsymbol{\delta}\right\|_z^2 \\ &= \left(\mu(z) - \nu\left(m(z)\right)\right)^2\left(1-|z|^2\right)^{-4}\delta_{ij}\delta_{k\ell}g^{ik}(z)g^{j\ell}(z) \\ &= \left(1 - \frac{\nu(m(z))}{\mu(z)}\right)^2.\end{aligned}$$

For every pair of μ,ν, we are now ready to define the distance function $d^R_{\mu,\nu}(\cdot,\cdot)$ on $\mathcal{D}$:

$$d^R_{\mu,\nu}(z_0,w_0) := \inf_{\substack{m\in M_{\mathcal{D}}\\ m(z_0)=w_0}} \int_{\Omega_{z_0,R}} \left|\mu(z) - \left(m^*\nu\right)(z)\right| d_{\mathrm{vol}_H}(z), \tag{3.5}$$

where $d_{\mathrm{vol}_H}(z) = (1-|z|^2)^{-2}\,dx^1 \wedge dx^2$ is the volume form for the hyperbolic disk. The integral in (3.5) can also be written in the following form, which makes its invariance more readily apparent:

$$\int_{\Omega_{z_0,R}} \left|1 - \frac{\nu(m(z))}{\mu(z)}\right| d_{\mathrm{vol}_{\mathcal{M}}}(z) = \int_{\Omega_{z_0,R}} \|\boldsymbol{g} - m^*\boldsymbol{h}\|_z d_{\mathrm{vol}_{\mathcal{M}}}(z), \tag{3.6}$$

where $d_{\mathrm{vol}_{\mathcal{M}}}(z) = \mu(z)(1-|z|^2)^{-2}\,dx^1 \wedge dx^2 = \sqrt{|g_{ij}|}\,dx^1 \wedge dx^2$ is the volume form of the first surface $\mathcal{M}$.

The next lemma shows that although the integration in (3.6) is carried out w.r.t. the volume of the first surface, this measure of distance is nevertheless symmetric:

Lemma 3.2. *If $m \in M_{\mathcal{D}}$ maps z_0 to w_0, $m(z_0) = w_0$, then*

$$\int_{\Omega_{z_0,R}} \left|\mu(z) - m^*\nu(z)\right| d_{\mathrm{vol}_H}(z) = \int_{\Omega_{w_0,R}} \left|m_*\mu(w) - \nu(w)\right| d_{\mathrm{vol}_H}(w).$$

Proof. By the pull-back formula (2.5), we have

$$\int_{\Omega_{z_0,R}} \left|\mu(z) - m^*\nu(z)\right| d_{\mathrm{vol}_H}(z) = \int_{\Omega_{z_0}} \left|\mu(z) - \nu\bigl(m(z)\bigr)\right| d_{\mathrm{vol}_H}(z).$$

Performing the change of coordinates $z = m^{-1}(w)$ in the integral on the right-hand side, we obtain

$$\int_{m(\Omega_{z_0,R})} \left|\mu\bigl(m^{-1}(w)\bigr) - \nu(w)\right| d_{\mathrm{vol}_H}(w),$$

where we have used that m^{-1} is an isometry and therefore preserves the volume element $d_{\mathrm{vol}_H}(w) = (1-|w|^2)^{-2}\,dy^1 \wedge dy^2$. By Lemma 3.1, $m(\Omega_{z_0,R}) = \Omega_{w_0,R}$; using the push-forward formula (2.6) then allows to conclude. □

Note that our point of view in defining our "distance" between z and w differs from the classical point of view in mass transportation: traditionally, $d(z,w)$ is some sort of *physical distance* between the points z and w; in our case $d^R_{\mu,\nu}(z,w)$ measures the dissimilarity of (neighborhoods of) z and w.

The next theorem lists some important properties of $d^R_{\mu,\nu}$; its proof is given in Appendix A.

Theorem 3.3. *The distance function $d^R_{\mu,\nu}(z,w)$ satisfies the following properties*:

(1) $d^R_{m_1^*\mu,m_2^*\nu}\bigl(m_1^{-1}(z_0), m_2^{-1}(w_0)\bigr) = d^R_{\mu,\nu}(z_0,w_0)$ *invariance under* (*well-defined*) *Möbius changes of coordinates*;

(2) $d^R_{\mu,\nu}(z_0,w_0) = d^R_{\nu,\mu}(w_0,z_0)$ *symmetry*;

(3) $d^R_{\mu,\nu}(z_0,w_0) \geqslant 0$ *non-negativity*;

(4) $d^R_{\mu,\nu}(z_0,w_0) = 0 \quad \Rightarrow \quad \Omega_{z_0,R}$ *in* $(\mathcal{D},\mu)$ *and* $\Omega_{w_0,R}$ *in* $(\mathcal{D},\nu)$ *are isometric*;

(5) $d^R_{m^*\nu,\nu}\bigl(m^{-1}(z_0), z_0\bigr) = 0$ *reflexivity*;

(6) $d^R_{\mu_1,\mu_3}(z_1,z_3) \leqslant d^R_{\mu_1,\mu_2}(z_1,z_2) + d^R_{\mu_2,\mu_3}(z_2,z_3)$ *triangle inequality.*

In addition, the function $d^R_{\mu,\nu} : \mathcal{D} \times \mathcal{D} \to \mathbb{R}$ is continuous. To show this, we first look a little more closely at the 1-parameter family of disk Möbius transformations that map one pre-assigned point $z_0 \in \mathcal{D}$ to another pre-assigned point $w_0 \in \mathcal{D}$.

Definition 3.4. For any pair of points $z_0, w_0 \in \mathcal{D}$, we denote by M_{D,z_0,w_0} the set of Möbius transformations that map z_0 to w_0.

This family of Möbius transformations is completely characterized by the following lemma:

Lemma 3.5. *For any $z_0, w_0 \in \mathcal{D}$, the set M_{D,z_0,w_0} constitutes a 1-parameter family of disk Möbius transformations, parametrized continuously over S^1 (the unit circle). More precisely, every $m \in M_{D,z_0,w_0}$ is of the form*

$$m(z) = \tau \frac{z-a}{1-\bar{a}z}, \quad \text{with } a = a(z_0, w_0, \sigma) := \frac{z_0 - w_0\bar{\sigma}}{1-\overline{z_0}w_0\bar{\sigma}} \text{ and}$$
$$\tau = \tau(z_0, w_0, \sigma) := \sigma \frac{1-\overline{z_0}w_0\bar{\sigma}}{1-z_0\overline{w_0}\sigma}, \tag{3.7}$$

where $\sigma \in S_1 := \{z \in \mathbb{C};\ |z| = 1\}$ can be chosen freely.

Proof. By (2.3), the disk Möbius transformations that map z_0 to 0 all have the form

$$m_{\psi,z_0}(z) = e^{\mathbf{i}\psi} \frac{z-z_0}{1-\overline{z_0}z}, \quad \text{the inverse of which is } m^{-1}_{\psi,z_0}(w) = e^{-\mathbf{i}\psi} \frac{w + e^{\mathbf{i}\psi} z_0}{1 + e^{-\mathbf{i}\psi}\overline{z_0}w},$$

where $\psi \in \mathbb{R}$ can be set arbitrarily. It follows that the elements of M_{D,z_0,w_0} are given by the family $m^{-1}_{\gamma,w_0} \circ m_{\psi,z_0}$, with $\psi, \gamma \in \mathbb{R}$. Working this out, one finds that these combinations of Möbius transformations take the form (3.7), with $\sigma = e^{\mathbf{i}(\psi-\gamma)}$. □

We shall denote by $m_{z_0,w_0,\sigma}$ the special disk Möbius transformation defined by (3.7). In view of our interest in $d^R_{\mu,\nu}$, we also define the auxiliary function

$$\Phi: \quad \mathcal{D} \times \mathcal{D} \times S_1 \to \mathbb{C},$$
$$(z_0, w_0, \sigma) \mapsto \int_{\Omega_{z_0,R}} \left|\mu(z) - \nu\big(m_{z_0,w_0,\sigma}(z)\big)\right| d_{\mathrm{vol}_H}(z).$$

This function has the following continuity properties, inherited from μ and ν:

Lemma 3.6.

- *For each fixed (z_0, w_0), the function $\Phi(z_0, w_0, \cdot)$ is continuous on S_1.*
- *For each fixed $\sigma \in S_1$, $\Phi(\cdot,\cdot,\sigma)$ is continuous on $\mathcal{D} \times \mathcal{D}$. Moreover, the family $(\Phi(\cdot,\cdot,\sigma))_{\sigma \in S_1}$ is equicontinuous.*

Proof. The proof of this lemma is given in Appendix A. □

Note that since S^1 is compact, Lemma 3.6 implies that the infimum in the definition of $d^R_{\mu,\nu}$ can be replaced by a minimum:

$$d^R_{\mu,\nu}(z_0, w_0) = \min_{m(z_0)=w_0} \int_{\Omega_{z_0,R}} \left|\mu(z) - \nu\big(m(z)\big)\right| d_{\mathrm{vol}_H}(z).$$

We have now all the building blocks to prove

Theorem 3.7. *If* μ *and* ν *are continuous from* $\mathcal{D}$ *to* $\mathbb{R}$, *then* $d^R_{\mu,\nu}(z,w)$ *is a continuous function on* $\mathcal{D}\times\mathcal{D}$.

Proof. Pick an arbitrary point $(z_0,w_0)\in\mathcal{D}\times\mathcal{D}$, and pick $\varepsilon>0$ arbitrarily small.

By Lemma 3.6, there exists a $\delta>0$ such that, for $|z_0'-z_0|<\delta$, $|w_0'-w_0|<\delta$, we have

$$\left|\Phi(z_0,w_0,\sigma)-\Phi\left(z_0',w_0',\sigma\right)\right|\leqslant\varepsilon,$$

uniformly in σ. Pick now arbitrary z_0', w_0' so that $|z_0-z_0'|$, $|w_0-w_0'|<\delta$.

Let $m_{z_0,w_0,\sigma}$, resp. $m_{z_0',w_0',\sigma'}$, be the minimizing Möbius transform in the definition of $d^R_{\mu,\nu}(z_0,w_0)$, resp. $d^R_{\mu,\nu}(z_0',w_0')$, i.e.

$$d^R_{\mu,\nu}(z_0,w_0)=\Phi(z_0,w_0,\sigma)\quad\text{and}\quad d^R_{\mu,\nu}\left(z_0',w_0'\right)=\Phi\left(z_0',w_0',\sigma'\right).$$

It then follows that

$$\begin{aligned}d^R_{\mu,\nu}(z_0,w_0)&=\min_\tau\Phi(z_0,w_0,\tau)\leqslant\Phi\left(z_0,w_0,\sigma'\right)\\&\leqslant\Phi\left(z_0',w_0',\sigma'\right)+\left|\Phi\left(z_0,w_0,\sigma'\right)-\Phi\left(z_0',w_0',\sigma'\right)\right|\\&=d^R_{\mu,\nu}\left(z_0',w_0'\right)+\left|\Phi\left(z_0,w_0,\sigma'\right)-\Phi\left(z_0',w_0',\sigma'\right)\right|\\&\leqslant d^R_{\mu,\nu}\left(z_0',w_0'\right)+\sup_{\omega\in S_1}\left|\Phi(z_0,w_0,\omega)-\Phi\left(z_0',w_0',\omega\right)\right|\\&\leqslant d^R_{\mu,\nu}\left(z_0',w_0'\right)+\varepsilon.\end{aligned}$$

Likewise $d^R_{\mu,\nu}(z_0',w_0')\leqslant d^R_{\mu,\nu}(z_0,w_0)+\varepsilon$, so that $|d^R_{\mu,\nu}(z_0,w_0)-d^R_{\mu,\nu}(z_0',w_0')|<\varepsilon$. □

The function $d^R_{\mu,\nu}$ can be extended to a uniformly continuous function on the closed disk, by using the following lemma, proved in Appendix A.

Lemma 3.8. *Let* $\{(z_k,w_k)\}_{k\geqslant1}\subset\mathcal{D}\times\mathcal{D}$ *be a sequence that converges, in the Euclidean norm, to some point in* $(z',w')\in\overline{\mathcal{D}}\times\overline{\mathcal{D}}\setminus\mathcal{D}\times\mathcal{D}$, *that is* $|z_k-z'|+|w_k-w'|\to0$, *as* $k\to\infty$. *Then,* $\lim_{k\to\infty}d^R_{\xi,\zeta}(z_k,w_k)$ *exists and depends only on the limit point* (z',w').

3.2. Incorporating $d^R_{\mu,\nu}(z,w)$ *into the transportation framework*

The next step in constructing the distance operator between surfaces is to incorporate the distance $d^R_{\mu,\nu}(z,w)$ defined in the previous subsection into the (generalized) Kantorovich transportation model:

$$T^R_d(\mu,\nu)=\inf_{\pi\in\Pi(\mu,\nu)}\int\limits_{\mathcal{D}\times\mathcal{D}}d^R_{\mu,\nu}(z,w)\,d\pi(z,w).\tag{3.8}$$

The main result is that this procedure (under some extra conditions) furnishes a *metric* between (disk-type) surfaces.

Theorem 3.9. *There exists $\pi^* \in \Pi(\mu,\nu)$ such that*

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

Proof. With a slight abuse of notation, we denote by μ the probability measure on $\mathcal{D}$ that is absolutely continuous with respect to the hyperbolic measure on $\mathcal{D}$, with density function equal to the continuous function μ on $\mathcal{D}$. (See Subsection 2.1.) We now define a probability measure $\overline{\mu}$ on $\overline{\mathcal{D}}$ by setting $\overline{\mu}(A) = \mu(A\cap\mathcal{D})$, for arbitrary Borel sets $A \subset \overline{\mathcal{D}}$; $\overline{\nu}$ is defined analogously. By the Riesz–Markov theorem, the space of probability measures $\mathcal{P}(\overline{\mathcal{D}}\times\overline{\mathcal{D}})$ can be viewed as a (closed) subset of the unit ball in $C(\overline{\mathcal{D}}\times\overline{\mathcal{D}})^*$. As such, both $\mathcal{P}(\overline{\mathcal{D}}\times\overline{\mathcal{D}})$ and its closed subset $\Pi(\overline{\mu},\overline{\nu})$ are weak∗-compact, by the Banach–Alaoglu theorem. Note that for each $\overline{\pi}\in\Pi(\overline{\mu},\overline{\nu})$, we have

$$\begin{aligned}\overline{\pi}(\mathcal{D}\times\mathcal{D}) &\geqslant \overline{\pi}(\overline{\mathcal{D}}\times\overline{\mathcal{D}}) - \big(\overline{\pi}\big(\overline{\mathcal{D}}\times[\overline{\mathcal{D}}\setminus\mathcal{D}]\big) + \overline{\pi}\big([\overline{\mathcal{D}}\setminus\mathcal{D}]\times\overline{\mathcal{D}}\big)\big)\\ &= 1 - \overline{\nu}(\overline{\mathcal{D}}\setminus\mathcal{D}) - \overline{\mu}(\overline{\mathcal{D}}\setminus\mathcal{D}) = 1,\end{aligned}$$

and thus $\overline{\pi}(\mathcal{D}\times\mathcal{D}) = 1$; the restriction π of each such $\overline{\pi}$ to the Borel sets contained in $\mathcal{D}\times\mathcal{D}$ is thus a probability measure on $\mathcal{D}\times\mathcal{D}$.

Since (the extension to $\overline{\mathcal{D}}\times\overline{\mathcal{D}}$ of) $d^R_{\mu,\nu}(\cdot\,,\cdot)$ is an element of $C(\overline{\mathcal{D}}\times\overline{\mathcal{D}})$ by Lemma 3.8, it follows that the evaluation

$$\overline{\pi}\mapsto\overline{\pi}(d_{\mu,\nu}) = \int_{\overline{\mathcal{D}}\times\overline{\mathcal{D}}} d^R_{\mu,\nu}(z,w)\,d\overline{\pi}(z,w) = \int_{\mathcal{D}\times\mathcal{D}} d^R_{\mu,\nu}(z,w)\,d\pi(z,w)$$

is weak∗-continuous on the weak∗-compact set $\Pi(\overline{\mu},\overline{\nu})$; it thus achieves its infimum in an element $\overline{\pi^*}$ of that set. As observed above, π^*, the restriction of $\overline{\pi^*}$ to the Borel sets contained in $\mathcal{D}\times\mathcal{D}$, is a probability measure on $\mathcal{D}\times\mathcal{D}$, and an element of $\Pi(\mu,\nu)$; this is the desired minimizer. □

Under rather mild conditions, the "standard" Kantorovich transportation (2.7) on a metric spaces (X,d) defines a metric on the space of probability measures on X. We will prove that our generalization defines a distance metric as well. More precisely, we shall prove first that

$$\mathbf{d}^R(\mathcal{M},\mathcal{N}) = T^R_d(\mu,\nu)$$

defines a semi-metric in the set of all disk-type surfaces. We shall restrict ourselves to surfaces that are sufficiently smooth to allow uniformization, so that they can be globally and conformally parametrized over the hyperbolic disk. Under some extra assumptions, we will prove that $\mathbf{d}^R$ is a metric, in the sense that $\mathbf{d}^R(\mathcal{M},\mathcal{N}) = 0$ implies that $\mathcal{M}$ and $\mathcal{N}$ are isometric.

For the semi-metric part we will adapt a proof given in [14] to our framework. In particular, we shall make use of the following "gluing lemma":

Lemma 3.10. *Let μ_1, μ_2, μ_3 be three probability measures on $\mathcal{D}$, and let $\pi_{12} \in \Pi(\mu_1, \mu_2)$, $\pi_{23} \in \Pi(\mu_2, \mu_3)$ be two transportation plans. Then there exists a probability measure π on $\mathcal{D} \times \mathcal{D} \times \mathcal{D}$ that has π_{12}, π_{23} as marginals, that is $\int_{z_3 \in \mathcal{D}} d\pi(z_1, z_2, z_3) = d\pi_{12}(z_1, z_2)$, and $\int_{z_1 \in \mathcal{D}} d\pi(z_1, z_2, z_3) = d\pi_{23}(z_2, z_3)$.*

This lemma will be used in the proof of the following:

Theorem 3.11. *For two disk-type surfaces $\mathcal{M} = (\mathcal{D}, \mu)$, $\mathcal{N} = (\mathcal{D}, \nu)$, let $\mathbf{d}^R(\mathcal{M}, \mathcal{N})$ be defined by*

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

Then $\mathbf{d}^R$ defines a semi-metric on the space of disk-type surfaces.

Proof. The symmetry of $d^R_{\mu,\nu}$ implies symmetry for T_d^R, by the following argument:

$$\begin{aligned}
T_d^R(\mu, \nu) &= \inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{D}\times\mathcal{D}} d^R_{\mu,\nu}(z, w)\, d\pi(z, w) \\
&= \inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{D}\times\mathcal{D}} d^R_{\nu,\mu}(w, z)\, d\pi(z, w) \\
&= \inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{D}\times\mathcal{D}} d^R_{\nu,\mu}(w, z)\, d\widetilde{\pi}(w, z), \quad \text{where we have set } \widetilde{\pi}(w, z) = \pi(z, w) \\
&= T_d^R(\nu, \mu) \quad \big(\text{use that } \pi \in \Pi(\mu, \nu) \Leftrightarrow \widetilde{\pi} \in \Pi(\nu, \mu)\big).
\end{aligned}$$

The non-negativity of $d^R_{\mu,\nu}(\cdot\,,\cdot)$ automatically implies $T_d^R(\mu, \nu) \geqslant 0$.

Next we show that, for any Möbius transformation m, $T_d^R(\mu, m_*\mu) = 0$. To see this, pick the transportation plan $\pi \in \Pi(\mu, m_*\mu)$ defined by

$$\int_{\mathcal{D}\times\mathcal{D}} f(z, w)\, d\pi(z, w) = \int_{\mathcal{D}} f\big(z, m(z)\big)\mu(z)\, d_{\mathrm{vol}_H}(z).$$

On the one hand $\pi \in \Pi(\mu, m_*\mu)$, since

$$\int_{A\times\mathcal{D}} d\pi(z, w) = \int_{A} \mu(z)\, d_{\mathrm{vol}_H}(z),$$

and

$$\begin{aligned}
\int_{\mathcal{D}\times B} d\pi(z, w) &= \int_{\mathcal{D}\times\mathcal{D}} \chi_B(w)\, d\pi(z, w) = \int_{\mathcal{D}} \chi_B\big(m(z)\big)\mu(z)\, d_{\mathrm{vol}_H}(z) \\
&= \int_{\mathcal{D}} \chi_B(w) m_*\mu(w)\, d_{\mathrm{vol}_H}(w),
\end{aligned}$$

where we used the change of variables $w = m(z)$ in the last step. Furthermore, $\pi(z, w)$ is concentrated on the graph of m, i.e. on $\{(z, m(z));\ z \in \mathcal{D}\} \subset \mathcal{D} \times \mathcal{D}$. Since $d^R_{\mu,m_*\mu}(z, m(z)) = 0$ for all $z \in \mathcal{D}$ we obtain therefore

$$T_d(\mu, m_*\mu) \leqslant \int\limits_{\mathcal{D}\times\mathcal{D}} d^R_{\mu,m_*\mu}(z, w)\, d\pi(z, w) = 0.$$

Finally, we prove the triangle inequality $T^R_d(\mu_1, \mu_3) \leqslant T^R_d(\mu_1, \mu_2) + T^R_d(\mu_2, \mu_3)$. To this end we follow the argument in the proof given in [14, p. 208]. This is where we invoke the gluing lemma stated above.

We start by picking arbitrary transportation plans $\pi_{12} \in \Pi(\mu_1, \mu_2)$ and $\pi_{23} \in \Pi(\mu_2, \mu_3)$. By Lemma 3.10 there exists a probability measure π on $\mathcal{D} \times \mathcal{D} \times \mathcal{D}$ with marginals π_{12} and π_{23}. Denote by π_{13} its third marginal, that is

$$\int\limits_{z_2\in\mathcal{D}} d\pi(z_1, z_2, z_3) = d\pi_{13}(z_1, z_3).$$

Then

$$\begin{aligned}
T^R_d(\mu_1, \mu_3) &\leqslant \int\limits_{\mathcal{D}\times\mathcal{D}} d^R_{\mu_1,\mu_3}(z_1, z_3)\, d\pi_{13}(z_1, z_3) \\
&= \int\limits_{\mathcal{D}\times\mathcal{D}\times\mathcal{D}} d^R_{\mu_1,\mu_3}(z_1, z_3)\, d\pi(z_1, z_2, z_3) \\
&\leqslant \int\limits_{\mathcal{D}\times\mathcal{D}\times\mathcal{D}} \big(d^R_{\mu_1,\mu_2}(z_1, z_2) + d^R_{\mu_2,\mu_3}(z_2, z_3)\big)\, d\pi(z_1, z_2, z_3) \\
&\leqslant \int\limits_{\mathcal{D}\times\mathcal{D}\times\mathcal{D}} d^R_{\mu_1,\mu_2}(z_1, z_2)\, d\pi(z_1, z_2, z_3) + \int\limits_{\mathcal{D}\times\mathcal{D}\times\mathcal{D}} d^R_{\mu_2,\mu_3}(z_2, z_3)\, d\pi(z_1, z_2, z_3) \\
&\leqslant \int\limits_{\mathcal{D}\times\mathcal{D}} d^R_{\mu_1,\mu_2}(z_1, z_2)\, d\pi_{12}(z_1, z_2) + \int\limits_{\mathcal{D}\times\mathcal{D}} d^R_{\mu_2,\mu_3}(z_2, z_3)\, d\pi_{23}(z_2, z_3),
\end{aligned}$$

where we used the triangle-inequality for $d^R_{\mu,\nu}$ listed in (Theorem 3.3). Since we can choose π_{12} and π_{23} to achieve arbitrary close values to the infimum in Eq. (3.8) the triangle inequality follows. □

To qualify as a metric rather than a semi-metric, $\mathbf{d}^R$ (or T^R_d) should be able to distinguish from each other any two surfaces (or measures) that are not "identical", that is isometric. To prove that they can do so, we need an extra assumption: we shall require that the surfaces we consider have no self-isometries. More precisely, we require that each surface $\mathcal{M}$ that we consider satisfies the following definition:

Definition 3.12. A disk-type surface $\mathcal{M}$ is said to be *singly ϱ-$_H$fittable* (where $\varrho \in \mathbb{R}$, $\varrho > 0$) if, for all $R > \varrho$, all $z_0 \in \mathcal{D}$, and all conformal factors obtained in uniformizations of the disk $\mathcal{M}$ there is no other Möbius transformation m other than the identity for which

$$\int_{\Omega_{z_0,R}} \left|\mu(z) - \mu\big(m(z)\big)\right| d_{\mathrm{vol}_H}(z) = 0.$$

Remark 3.13. This definition can also be read as follows: $\mathcal{M}$ is singly ϱ-$_H$fittable if and only if, for all $R > \varrho$, any two conformal factors μ_1 and μ_2 for $\mathcal{M}$ satisfy:

(1) For all $z \in \mathcal{D}$ there exists a unique minimum to the function $w \mapsto d^R_{\mu_1,\mu_2}(z, w)$.
(2) For all pairs $(z, w) \in \mathcal{D} \times \mathcal{D}$ that achieve this minimum there exists a unique Möbius transformation for which the integral in (3.5) vanishes (with μ_1 in the role of μ, and μ_2 in that of ν).

Note that in order to ensure that the conditions in the definition hold for all conformal factors, it is sufficient to require that it holds for the conformal factor associated to just one uniformization.

Essentially, this definition requires that, from some sufficiently large (hyperbolic) scale onwards, there are no isometric pieces within $(\mathcal{D}, \mu)$ (or $(\mathcal{D}, \nu)$).

We are ready to prove the last remaining part of the main result of this subsection. We start with a lemma.

Lemma 3.14. *Let $\pi \in \Pi(\mu, \nu)$ be such that $\int_{\mathcal{D}\times\mathcal{D}} d^R_{\mu,\nu}(z, w)\, d\pi(z, w) = 0$. Then, for all $z_0 \in \mathcal{D}$ and $\delta > 0$, there exists at least one point $z \in \Omega_{z_0,\delta}$ such that $d^R_{\mu,\nu}(z, w) = 0$ for some $w \in \mathcal{D}$.*

Proof. By contradiction: assume that there exists a disk $\Omega_{z_0,\delta}$ such that $d^R_{\mu,\nu}(z, w) > 0$ for all $z \in \Omega_{z_0,\delta}$ and all $w \in \mathcal{D}$. Since

$$\int_{\Omega(z_0,\delta)\times\mathcal{D}} d\pi(z, w) = \int_{\Omega(z_0,\delta)} \mu(z)\, d_{\mathrm{vol}_H}(z) > 0,$$

the set $\Omega(z_0, \delta) \times \mathcal{D}$ contains some of the support of π. It follows that

$$\int_{\Omega(z_0,\delta)\times\mathcal{D}} d^R_{\mu,\nu}(z, w)\, d\pi(z, w) > 0,$$

which contradicts

$$\int_{\Omega(z_0,\delta)\times\mathcal{D}} d^R_{\mu,\nu}(z, w)\, d\pi(z, w) \leqslant \int_{\mathcal{D}\times\mathcal{D}} d^R_{\mu,\nu}(z, w)\, d\pi(z, w) = 0. \qquad \square$$

Theorem 3.15. *Suppose that $\mathcal{M}$ and $\mathcal{N}$ are two surfaces that are singly ϱ-$_H$fittable. If $\mathbf{d}^R(\mathcal{M}, \mathcal{N}) = 0$ for some $R > \varrho$, then there exists a Möbius transformation $m \in M_{\mathcal{D}}$ that is a*

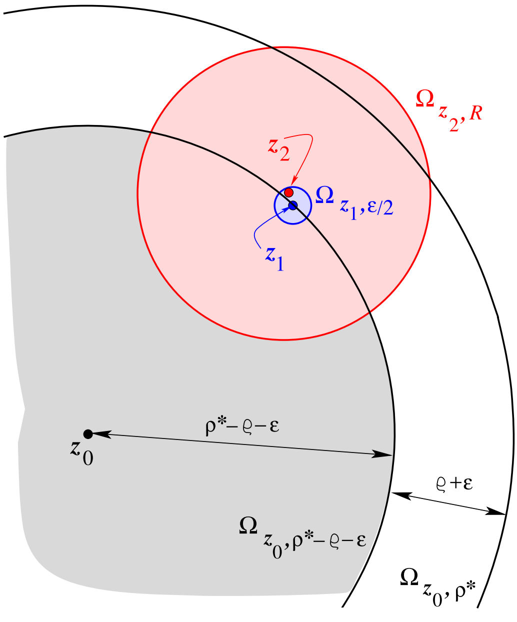


Fig. 1. Illustration of the proof of Theorem 3.15.

global isometry between $\mathcal{M} = (\mathcal{D}, \mu)$ *and* $\mathcal{N} = (\mathcal{D}, \nu)$ (*where* μ *and* ν *are conformal factors of* $\mathcal{M}$ *and* $\mathcal{N}$*, respectively*).

Proof. When $\mathbf{d}^R(\mathcal{M}, \mathcal{N}) = 0$, there exists $\pi \in \Pi(\mu, \nu)$ such that

$$\int_{\mathcal{D}\times\mathcal{D}} d^R_{\mu,\nu}(z, w)\, d\pi(z, w) = 0.$$

Next, pick an arbitrary point $z_0 \in \mathcal{D}$ such that, for some $w_0 \in \mathcal{D}$, we have $d^R_{\mu,\nu}(z_0, w_0) = 0$. (The existence of such a pair is guaranteed by Lemma 3.14.) This implies that there exists a unique Möbius transformation $m_0 \in M_{\mathcal{D}}$ that takes z_0 to w_0 and that satisfies $\nu(m_0(z)) = \mu(z)$ for all $z \in \Omega_{z_0,R}$. We define

$$\rho^* = \sup\bigl\{\rho;\ d^{\rho}_{\mu,\nu}(z_0, w_0) = 0\bigr\};$$

clearly $\rho^* \geqslant R$. The theorem will be proved if we show that $\rho^* = \infty$. We shall do this by contradiction.

Assume $\rho^* < \infty$. Consider Ω_{z_0,ρ^*}, the hyperbolic disk around z_0 of radius ρ^*. (See Fig. 1 for illustration.) Set $\varepsilon = (R - \varrho)/2$, and consider the points on the hyperbolic circle $C = \partial\Omega_{z_0,\rho^*-\varrho-\varepsilon}$. For every $z_1 \in C$, consider the hyperbolic disk $\Omega_{z_1,\varepsilon/2}$; by Lemma 3.14 there exists a point z_2 in this disk and a corresponding point $w_2 \in \mathcal{D}$ such that $d^R_{\mu,\nu}(z_2, w_2) = 0$, i.e. such that

$$\int_{\Omega_{z_2,R}} \bigl|\mu(z) - m'^*\nu(z)\bigr|\, d_{\mathrm{vol}_H}(z) = 0$$

for some Möbius transformation m' that maps z_2 to w_2; in particular, we have that

$$\mu(z) = \nu\big(m'(z)\big) \quad \text{for all } z \in \Omega_{z_2,R}. \tag{3.9}$$

The hyperbolic distance from z_2 to $\partial\Omega_{z_0,\rho^*}$ is at least $\varrho + \varepsilon/2$. It follows that the hyperbolic disk $\Omega_{z_2,\varrho+\varepsilon/4}$ is completely contained in Ω_{z_0,ρ^*}; since $\mu(z) = \nu(m_0(z))$ for all $z \in \Omega_{z_0,\rho^*}$, this must therefore hold, in particular, for all $z \in \Omega_{z_2,\varrho+\varepsilon/4}$. Since $\Omega_{z_2,\varrho+\varepsilon/4} \subset \Omega_{z_2,R}$, we also have $\mu(z) = \nu(m'(z))$ for all $z \in \Omega_{z_2,\varrho+\varepsilon/4}$, by (3.9). This implies $\nu(w) = \nu(m_0 \circ (m')^{-1}(w))$ for all $w \in \Omega_{w_2,\varrho+\varepsilon/4}$. Because $\mathcal{N}$ is singly ϱ-$_H$fittable, it follows that $m_0 \circ (m')^{-1}$ must be the identity, or $m_0 = m'$. Combining this with (3.9), we have thus shown that $\mu(z) = \nu(m_0(z))$ for all $z \in \Omega_{z_2,R}$.

Since the distance between z_2 and z_1 is at most $\varepsilon/2$, we also have

$$\Omega_{z_2,R} \supset \Omega_{z_1,R-\varepsilon/2} = \Omega_{z_1,\varrho+3\varepsilon/2}.$$

This implies that if we select such a point $z_2(z_1)$ for each $z_1 \in C$, then $\Omega_{z_0,\rho^*-\varrho-\varepsilon} \cup (\bigcup_{z_1\in C} \Omega_{z_2(z_1),R})$ covers the open disk $\Omega_{z_0,\rho^*+\varepsilon/2}$. By our earlier argument, $\mu(z) = \nu(m_0(z))$ for all z in each of the $\Omega_{z_2(z_1),R}$; since the same is true on $\Omega_{z_0,\rho^*-\varrho-\varepsilon}$, it follows that $\mu(z) = \nu(m_0(z))$ for all z in $\Omega_{z_0,\rho^*+\varepsilon/2}$. This contradicts the definition of ρ^* as the supremum of all radii for which this was true; it follows that our initial assumption, that ρ^* is finite, cannot be true, completing the proof. □

For $(\mathcal{D}, \mu)$ to be singly ϱ-$_H$fittable, no two hyperbolic disks $\Omega_{z,R}$, $\Omega_{w,R}$ (where w can equal z) can be isometric via a Möbius transformation m, if $R > \varrho$, except if $m = Id$. However, if z is close (in the Euclidean sense) to the boundary of $\mathcal{D}$, the hyperbolic disk $\Omega_{z,R}$ is very small in the Euclidean sense, and corresponds to a very small piece (near the boundary) of $\mathcal{M}$. This means that single ϱ-$_H$fittability imposes restrictions in increasingly small scales near the boundary of $\mathcal{M}$; from a practical point of view, this is hard to check, and in many applications, the behavior of $\mathcal{M}$ close to its boundary is irrelevant. For this reason, we also formulate the following relaxation of the results above.

Definition 3.16. A surface $\mathcal{M}$ is said to be *singly A-$_\mathcal{M}$fittable* (where $A > 0$) if there are no patches (i.e. open, path-connected sets) in $\mathcal{M}$ of area larger than A that are isometric, with respect to the metric on $\mathcal{M}$.

If a surface is singly A-$_\mathcal{M}$fittable, then it is obviously also A'-$_\mathcal{M}$fittable for all $A' \geqslant A$; the condition of being A-$_\mathcal{M}$fittable becomes more restrictive as A decreases. The following theorem states that two singly A-$_\mathcal{M}$fittable surfaces at zero $\mathbf{d}^R$-distance from each other must necessarily be isometric, up to some small boundary layer.

Theorem 3.17. *Consider two surfaces $\mathcal{M}$ and $\mathcal{N}$, with corresponding conformal factors μ and ν on $\mathcal{D}$, and suppose $\mathbf{d}^R(\mathcal{M},\mathcal{N}) = 0$ for some $R > 0$. Then the following holds: for arbitrarily large $\rho > 0$, there exist a Möbius transformation $m \in M_D$ and a value $A > 0$ such that if $\mathcal{M}$ and $\mathcal{N}$ are singly A-$_\mathcal{M}$fittable then $\mu(m(z)) = \nu(z)$, for all $z \in \Omega_{0,\rho}$.*

Proof. Part of the proof follows the same lines as for Theorem 3.15. We highlight here only the new elements needed for this proof.

First, note that, for arbitrary $r > 0$ and $z_0 \in \mathcal{D}$,

$$\mathrm{vol}_{\mathcal{M}}(\Omega_{z_0,r}) = \int\limits_{\Omega_{z_0,r}} \mu(z)\, d_{\mathrm{vol}_H}(z) \geqslant \mathrm{vol}_H(\Omega_{z_0,r})\Big[\min_{z\in\Omega_{z_0,r}} \mu(z)\Big]$$

$$= \mathrm{vol}_H(\Omega_{0,r})\Big[\min_{z\in\Omega_{z_0,r}} \mu(z)\Big]. \tag{3.10}$$

This motivates the definition of the sets $\mathcal{O}_{A,r}$,

$$\mathcal{O}_{A,r} = \left\{ z \in \mathcal{D} \;\Big|\; \min_{z'\in\Omega_{z,r}} \mu(z') > \frac{A}{\mathrm{vol}_H(0, \Omega_{0,r})} \right\}; \tag{3.11}$$

$A > 0$ is still arbitrary at this point; its value will be set below.

Now pick $r < R$, and set $\varepsilon = (R - r)/2$. Note that if $z \in \mathcal{O}_{A,r}$, then $\mathrm{vol}_{\mathcal{M}}(\Omega_{z,R}) \geqslant \mathrm{vol}_{\mathcal{M}}(\Omega_{z,r}) > A$.

Since μ is bounded below by a strictly positive constant on each $\Omega_{0,\rho'}$, we can pick, for arbitrarily large ρ, $A > 0$ such that $\Omega_{0,\rho} \subset \mathcal{O}_{A,r}$; for this it suffices that A exceed a threshold depending on ρ and r. (Since $\mu(z) \to 0$ as z approaches the boundary of $\mathcal{D}$ in Euclidean norm, we expect this threshold to tend towards 0 as $\rho \to \infty$.) We assume that $\Omega_{0,\rho} \subset \mathcal{O}_{A,r}$ in what follows.

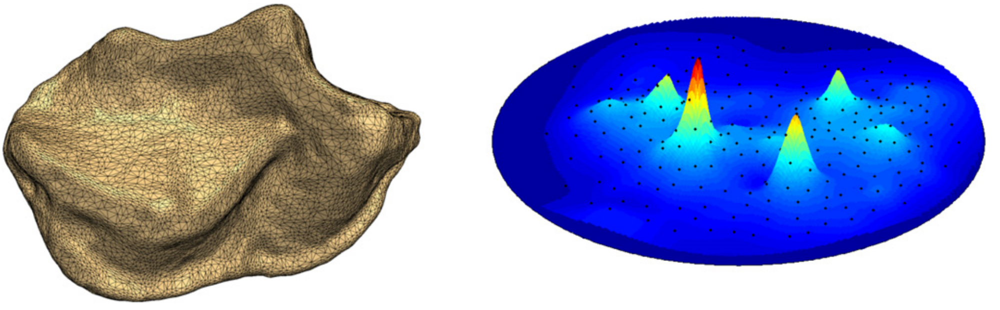
Similar to the proof of Theorem 3.15, we invoke Lemma 3.14 to infer the existence of z_0, w_0 such that $z_0 \in \Omega_{0,\varepsilon/2}$ and $d^R_{\mu,\nu}(z_0, w_0) = 0$. We denote

$$\rho^* = \sup\{r';\ d^{r'}_{\mu,\nu}(z_0, w_0) = 0\};$$

as before, there exists a Möbius transformation m such that $\nu(m(z)) = \mu(z)$ for all z in Ω_{z_0,ρ^*}. To complete our proof it therefore suffices to show that $\rho^* \geqslant \rho + \varepsilon/2$, since $\Omega_{0,\rho} \subset \Omega_{z_0,\rho+\varepsilon/2}$.

Suppose the opposite is true, i.e. $\rho^* < \rho + \varepsilon/2$. By the same arguments as in the proof of Theorem 3.15, there exists, for each $z_1 \in \partial\Omega_{z_0,\rho^*-r-\varepsilon}$, a point $z_2 \in \Omega_{z_1,\varepsilon/2}$ such that $d^R_{\mu,\nu}(z_2, w_2) = 0$ for some w_2. Since the hyperbolic distance between z_2 and 0 is bounded above by $\varepsilon/2 + \rho^* - r - \varepsilon + \varepsilon/2 < \rho - r + \varepsilon/2 < \rho$, $z_2 \in \Omega_{0,\rho} \subset \mathcal{O}_{A,r}$, so that $\mathrm{vol}_{\mathcal{M}}(\Omega_{z_2,R}) > A$. It then follows from the conditions on $\mathcal{M}$ and $\mathcal{N}$ that $\nu(m(z)) = \mu(z)$ for all z in $\Omega_{z_0,\rho^*} \cup \Omega_{z_2,R} \supset \Omega_{z_0,\rho^*} \cup \Omega_{z_1,r+3\varepsilon/2}$. Repeating the argument for all $z_1 \in \partial\Omega_{z_0,\rho^*-r-\varepsilon}$ shows that $\nu(m(z)) = \mu(z)$ can be extended to all $z \in \Omega_{z_0,\rho^*+\varepsilon/2}$, leading to a contradiction that completes the proof. □

So far, we have dealt exclusively with disk-type surfaces. The approach presented here can also be used for other surfaces, however. In order to apply this approach to sphere-type (genus zero) surfaces, for instance, we would need to change only one component in the construction, namely how to define the neighborhoods $\Omega_{z_0,R}$ in a Möbius-invariant way. Since there exists no Möbius-invariant distance function on the sphere, we cannot define the neighborhoods $\Omega_{z_0,R}$ as disks with respect to such an invariant distance. We thus need a different criterium to pick, among all the circles centered at a point $z_0 \in \mathcal{M}$, the one circle that shall delimit $\Omega_{z_0,R}$. Since the family of circles *is* invariant under (general) Möbius transformations, it suffices to pick a criterium that is itself invariant as well. For our applications, we pick $\widetilde{\Omega}_{z_0,A}$ to be the interior of the circle around z_0 that has the smallest circumference among all such circles with area (or



The discrete representation of a surface (mesh) Conformal density over the unit disk

Fig. 2. A mammalian tooth discrete surface mesh, and its approximated conformal factor over the unit disk.

volume) A (where $A \in (0, 1)$). In the generic case this procedure defines a unique neighborhood that can be used in the same way as $\Omega_{z_0,R}$ up till now.

For higher genus (but still homeomorphic) surfaces $\mathcal{M}, \mathcal{N}$ it is natural to use the universal coverings $\widetilde{\mathcal{M}}, \widetilde{\mathcal{N}}$, respectively. For genus greater than one, we could use again the disk-type construction. The fix here (to avoid infinite volume of the flattened surface via the universal covering) could be to restrict the compared $z_0 \in \mathcal{D}$ to *one copy* of $\mathcal{M}$ in the hyperbolic disk (and similarly $w_0 \in \mathcal{D}$ in $\mathcal{N}$). Treating the genus one case can be done similarly with similarity transformations in $\mathbb{C}$.

4. Discretization and implementation

Several steps are needed to transform the theoretical framework of the preceding sections into an algorithm, as described in detail in [8]. In a nutshell, the procedure requires three approximation steps: 1) approximating the smooth surfaces with a discrete mesh, 2) using discrete conformal theory to construct a discrete analog of uniformization for meshes, and 3) reducing the discrete optimization problem (resulting from replacing μ, ν in Eq. (3.8) by their discrete versions supported on a finite set of points) to a linear program. If an equal number of discrete point masses is chosen for the discrete measure on each of the two surfaces, and all of them are given equal weight, the corresponding search for the optimal bistochastic matrix automatically produces a minimizer that is a *permutation*. This means that the minimizer defines a map from (the discretized version of) one surface to (the discretized version of) the other.

It follows that the surface distance given in this paper does indeed lead to a computationally efficient approach, both for finding the best similarity distance and for identifying the best correspondence between two (disk-type) surfaces. Fig. 2 shows an example of a discrete surface and the corresponding approximate conformal density visualized as a graph over the unit disk.

Efficient computation of a distance between surfaces is important for many applications. As an example, Fig. 3 shows an application of our approach to the characterization of mammals by the surfaces of their molars, comparing high resolution scans of the masticating surfaces of molars of several lemurs (small primates living in Madagascar). The figure shows an embedding of eight molars, coming from individuals in four different species (indicated by color). The embedding is based on the pairwise distance matrix $(\mathbf{d}^R(\mathcal{M}_i, \mathcal{M}_j))$, and it clearly agrees with the clustering by species, as communicated to us by the biologists from whom we obtained the data sets.

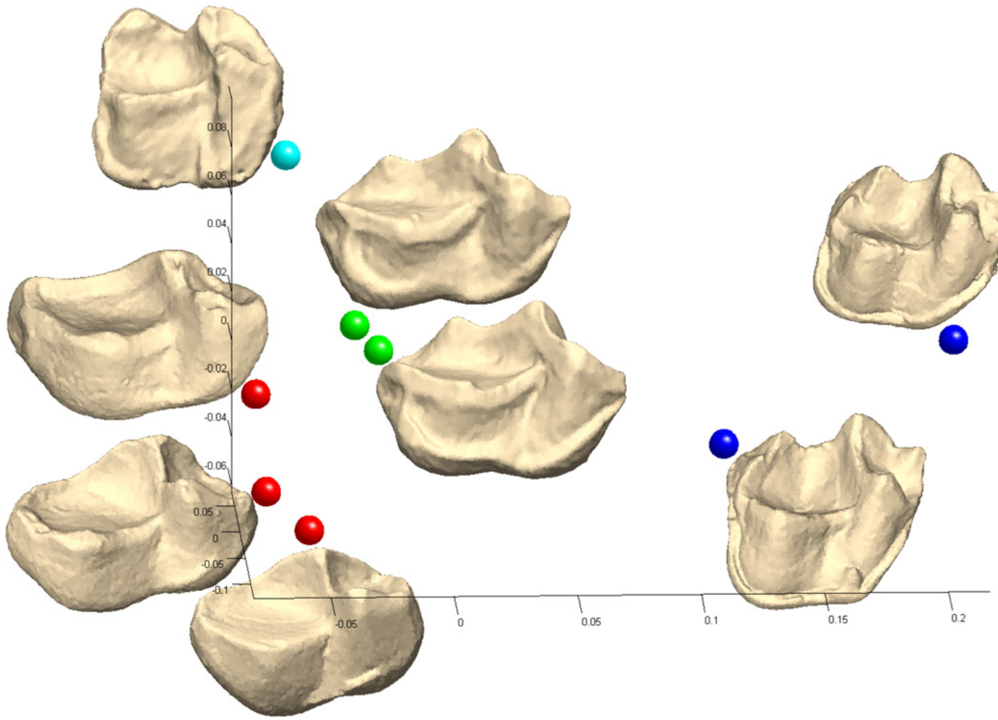


Fig. 3. Embedding of the distance graph of eight teeth models using multi-dimensional scaling. Different colors represent different lemur species. The graph suggests that the geometry of the teeth might suffice to classify species. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Acknowledgments

The authors would like to thank Cédric Villani and Thomas Funkhouser for valuable discussions, and Jesus Puente for helping with the implementation. We are grateful to Jukka Jernvall, Stephen King, and Doug Boyer for providing us with the tooth data sets, and for many interesting comments. ID gratefully acknowledges (partial) support for this work by NSF grant DMS-0914892, and by an AFOSR Complex Networks grant; Y.L. thanks the Rothschild foundation for postdoctoral fellowship support.

Appendix A

This appendix contains some technical proofs of lemmas and theorems stated in Sections 3 and 4. We start by proving that (3.1) does indeed define a distance metric on the family of $\overline{\mu} := \{m_*\mu;\ m \in M_{\mathcal{D}}\}$, where the μ are (smooth) conformal factors on $\mathcal{D}$, as obtained in Sections 2 and 3. We first state a more general lemma:

Lemma A.1. *Let X, d be a metric space, G a group, and $T : g \mapsto T_g$ a representation of G into the isometries of X, d, with d invariant under the action of the group G, i.e. $d(T_g x, T_g y) = d(x, y)$, for all $x, y \in X$ and all $g \in G$. Define $\mathcal{C}$ to be the collection of orbits of the representation of G, i.e. the elements of $\mathcal{C}$ take of the form $\{T_g x;\ g \in G\}$, for some $x \in X$. Define $\widetilde{d}$ on $\mathcal{C} \times \mathcal{C}$ by $\widetilde{d}(c_1, c_2) = \inf_{x_1 \in c_1, x_2 \in c_2} d(x_1, x_2)$. Then $\widetilde{d}$ defines a semi-metric on $\mathcal{C}$.*

Proof. It is obvious that $\widetilde{d}(c_1, c_2) \geqslant 0$ for all c_1, c_2 in $\mathcal{C}$; thus only the triangle inequality needs to be established.

Since an element c_1 of $\mathcal{C}$ can always be written as $c_1 = \{T_g x;\ g \in G\}$, where x is an arbitrary element of X, we obtain, for arbitrary c_1, c_2, c_3 in $\mathcal{C}$,

$$
\begin{aligned}
\widetilde{d}(c_1, c_3) &= \inf_{g, g' \in G} d(T_g x, T_{g'} z) \quad \text{where } x \in c_1, z \in c_3 \text{ are arbitrary} \\
&\leqslant d(T_{g_1} x, T_{g_3} z) \quad \text{for all } g_1, g_3 \in G \\
&\leqslant d(T_{g_1} x, T_{g_2} y) + d(T_{g_2} y, T_{g_3} z) \quad \text{for all } g_1, g_2, g_3 \in G \text{ and all } y \in X \\
&= d(T_{g_1} x, T_{g_2} y) + d(T_{g_2'} y, T_{g_2'(g_2)^{-1} g_3} z) \quad \text{for all } g_1, g_2, g_2', g_3 \in G \text{ and all } y \in X.
\end{aligned}
$$

When g_2', g_2 in G are kept fixed, the group elements $g_2'(g_2)^{-1} g_3$ run through all of G as g_3 varies over G. By taking the infimum over the choices of $g_1, g_2, g_2', g_3 \in G$ in the last expression, we thus obtain

$$
\widetilde{d}(c_1, c_3) \leqslant \widetilde{d}(c_1, c_2) + \widetilde{d}(c_2, c_3),
$$

where $c_2 := \{T_g y;\ g \in G\}$. Since $y \in X$ is arbitrary, this proves the triangle inequality in $\mathcal{C}$ for all three-tuples in $\mathcal{C}$. □

Note that one can use the invariance of d under the action of the group on X to define $\widetilde{d}(c_1, c_2)$ via a single minimization (instead of two): for $x \in c_1$, $y \in c_2$,

$$
\widetilde{d}(c_1, c_2) = \inf_{g, g' \in G} d(T_g x, T_{g'} y) = \inf_{g, g' \in G} d(x, T_{g^{-1} g'} y) = \inf_{g'' \in G} d(x, T_{g''} y).
$$

To apply this to (3.1), we choose X to be the set of non-negative C^1-functions on $\mathcal{D}$ that have integral 1 with respect to the hyperbolic area measure on $\mathcal{D}$, and d the Kantorovich mass transport distance d between them, with the "work" measured in terms of the hyperbolic distance metric d^{H} on $\mathcal{D}$:

$$
\mathrm{d}(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathcal{D} \times \mathcal{D}} d^{\mathrm{H}}(z, w)\, d\pi(z, w).
$$

The group G is here given by $M_{\mathcal{D}}$, and the action of G on X by push-forward: $T_m(\mu) = m_* \mu$.

Since $d^{\mathrm{H}}(z, w)$ is unbounded on $\mathcal{D}$, it may not be immediately obvious that $\mathcal{D}(\mu, \nu)$ is necessarily finite. The following argument shows that it is. Since we minimize over $m \in M_{\mathcal{D}}$ and $\pi \in \Pi(m_* \mu, \nu)$, it suffices to show finiteness for a specific choice of m and π. Take $m = Id$, and $\pi = \mu \times \nu$. Then we have:

$$
\mathrm{d}(\mu, \nu) \leqslant \int_{\mathcal{D} \times \mathcal{D}} d^{\mathrm{H}}(z, w)\, d\pi(z, w)
$$

$$\leqslant \int\limits_{\mathcal{D}\times\mathcal{D}} d^{\mathrm{H}}(z,0)\,d\pi(z,w) + \int\limits_{\mathcal{D}\times\mathcal{D}} d^{\mathrm{H}}(w,0)\,d\pi(z,w)$$

$$= \int\limits_{\mathcal{D}} d^{\mathrm{H}}(z,0)\mu(z)\,d_{\mathrm{vol}_H}(z) + \int\limits_{\mathcal{D}} d^{\mathrm{H}}(w,0)\nu(w)\,d_{\mathrm{vol}_H}(w).$$

Since $\mu(z)\,d_{\mathrm{vol}_H}(z) = \widetilde{\mu}(z)\,dx^1 \wedge dx^2$, and $\widetilde{\mu}(z) \leqslant C_{\widetilde{\mu}}$ for some constant $C_{\widetilde{\mu}} > 0$ for all $z \in \mathcal{D}$, we just need to show $\int_{\mathcal{D}} d^{\mathrm{H}}(z,0)\,dx^1 \wedge dx^2 < \infty$ (the same argument will apply to ν). But $d^{\mathrm{H}}(z,0) = \ln[\frac{1+|z|}{1-|z|}]$, and therefore the integral can easily be seen to converge by changing to polar coordinates.

To apply Lemma A.1, we first need to establish that:

Lemma A.2. $\mathrm{d}(\mu,\nu) = \mathrm{d}(m_*\mu, m_*\nu)$, *for all* μ, ν *in* X, *and all* m *in* $M_{\mathcal{D}}$.

Proof. We first rewrite $\mathrm{d}(m_*\mu, m_*\nu)$ in a different way. For each $\pi \in \Pi(\mu,\nu)$, we define the probability measure $m_*\pi$ on $\mathcal{D}\times\mathcal{D}$ by $m_*\pi(E) = \pi(\{(m^{-1}z, m^{-1}w);\ (z,w) \in E\})$. It is straightforward to check that $m_*\pi(A\times\mathcal{D}) = m_*\mu(A)$ and $m_*\pi(\mathcal{D}\times B) = m_*\nu(B)$ for all Borel sets $A, B \subset \mathcal{D}$; thus $m_*\pi \in \Pi(m_*\mu, m_*\nu)$. One can analogously define $m^*\pi$; again it is straightforward that $m^*m_*\pi = \pi$. It follows that $\Pi(m_*\mu, m_*\nu)$ is exactly equal to $\{m_*\pi;\ \pi \in \Pi(\mu,\nu)\}$.

Consequently, using the invariance $d^{\mathrm{H}}(mz, mw) = d^{\mathrm{H}}(z,w)$, we obtain

$$\mathrm{d}(m_*\mu, m_*\nu) = \inf_{\pi\in\Pi(\mu,\nu)} \int\limits_{\mathcal{D}\times\mathcal{D}} d^{\mathrm{H}}(z,w)\,d(m_*\pi)(z,w)$$

$$= \inf_{\pi\in\Pi(\mu,\nu)} \int\limits_{\mathcal{D}\times\mathcal{D}} d^{\mathrm{H}}(mu,mv)\,d\pi(u,v)$$

$$= \inf_{\pi\in\Pi(\mu,\nu)} \int\limits_{\mathcal{D}\times\mathcal{D}} d^{\mathrm{H}}(u,v)\,d\pi(u,v) = \mathrm{d}(\mu,\nu). \qquad \square$$

It follows that we can indeed apply Lemma A.1, and that (3.1) defines a semi-metric on the equivalence classes of conformal factors, where two conformal factors are viewed as equivalent if one can be obtained from the other by pushing it forward (or backward) through a Möbius transformation.

It turns out that in this case, the infimum over the choices $m \in M_{\mathcal{D}}$ is in fact always achieved (and is thus a minimum):

Lemma A.3. *Let* μ *and* ν *be conformal factors obtained by uniformizing two smooth disk-type surfaces, with* $\mathfrak{D}(\mu,\nu) < \infty$ *defined as in* (3.1). *Then there exists a Möbius transformation* $m \in M_{\mathcal{D}}$ *such that* $\mathfrak{D}(\mu,\nu) = \mathrm{d}(m_*\mu,\nu)$.

Proof. Consider two arbitrary (but fixed) conformal factors μ and ν on $\mathcal{D}$. There exists a sequence $(m_n)_{n\in\mathbb{N}}$ such that $\inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathcal{D}\times\mathcal{D}} d^{\mathrm{H}}(m_n(z),w)\,d\pi(z,w) \to \mathfrak{D}(\mu,\nu)$ as $n\to\infty$. Each of these m_n can be written in the form given by (2.3), with corresponding $a_n \in \mathcal{D}$, and $e^{\mathbf{i}\theta_n} \in \mathbb{T} := \{z \in \mathbb{C};\ |z| = 1\}$. By passing to a subsequence if necessary, we can assume, without loss of generality, that the sequences $(a_n)_{n\in\mathbb{N}}$ and $(e^{\mathbf{i}\theta_n})_{n\in\mathbb{N}}$ converge in $\overline{\mathcal{D}}$ (the closure of $\mathcal{D}$) and $\mathbb{T}$, respectively, to limits we denote by $\bar{a}$ and $e^{\mathbf{i}\bar{\theta}}$.

If $\bar{a}$ lies in the open disk $\mathcal{D}$, then it defines, together with $e^{\mathbf{i}\bar{\theta}}$, a corresponding $\overline{m} \in M_{\mathcal{D}}$. We then have, for all z, w in $\mathcal{D}$, $\lim_{n\to\infty} d^{\mathrm{H}}(m_n(z), w) = d^{\mathrm{H}}(\overline{m}(z), w)$. On the other hand, for sufficiently large n we have

$$\begin{aligned} d^{\mathrm{H}}\big(m_n(z), w\big) &\leqslant d^{\mathrm{H}}\big(m_n(z), 0\big) + d^{\mathrm{H}}(0, w) \\ &= d^{\mathrm{H}}(z, a_n) + d^{\mathrm{H}}(0, w) \\ &\leqslant d^{\mathrm{H}}(z, \bar{a}) + 1 + d^{\mathrm{H}}(0, w) \\ &\leqslant d^{\mathrm{H}}(z, 0) + d^{\mathrm{H}}(0, \bar{a}) + 1 + d^{\mathrm{H}}(0, w), \end{aligned}$$

where we have used the invariance of d^{H} under Möbius transformations and $m_n(a_n) = 0$ in the second line, and where we assume n sufficiently large to ensure $d^{\mathrm{H}}(a_n, \bar{a}) \leqslant 1$ in the third.

Therefore $d^{\mathrm{H}}(m_n(z), w)$ is bounded, uniformly in n, by a function that is absolutely integrable with respect to π (by the argument used just before the statement of Lemma A.2); the dominated convergence theorem then implies that

$$\begin{aligned} \mathfrak{D}(\mu, \nu) &= \lim_{n\to\infty} \int\limits_{\mathcal{D}\times\mathcal{D}} d^{\mathrm{H}}\big(m_n(z), w\big)\, d\pi(z, w) \\ &= \int\limits_{\mathcal{D}\times\mathcal{D}} d^{\mathrm{H}}\big(\overline{m}(z), w\big)\, d\pi(z, w), \end{aligned}$$

so that we are done for the case where $\bar{a} \in \mathcal{D}$.

It remains to discuss the case where $\bar{a} \in \overline{\mathcal{D}} \setminus \mathcal{D} = \mathbb{T}$, i.e. $|a| = 1$. The proof will be complete if we show that this is impossible; we will establish this by contradiction.

From now on, we suppose that $|\bar{a}| = 1$. By the integrability of μ and ν, we can find an increasing sequence of $\rho_n < 1$ such that $\mu(\{z;\ 1 > |z| > \rho_n\}) < 1/n$ and $\nu(\{z;\ 1 > |z| > \rho_n\}) < 1/n$. It is easy to check that

$$\text{for } |a| > R \text{ and } |z| < \rho < R\text{:} \quad \left|\frac{z - a}{1 - a^* z}\right| > \frac{R - \rho}{1 - R\rho}.$$

This lower bound tends to 1 as R tends to 1, regardless of the value of $\rho < 1$. It follows that there exist $R_n < 1$ so that

$$\inf_{|z|<\rho_n} \left|\frac{z - a}{1 - a^* z}\right| > (n + \rho_n)/(n + 1), \quad \text{for all } a \text{ with } |a| > R_n.$$

Because $|\bar{a}| = 1$, we can find a $k_n \in \mathbb{N}$ such that $|a_k| > R_n$ for all $k > k_n$; consequently $|m_k(z)| = |z - a_k|/|1 - (a_k)^* z| > (n + \rho_n)/(n + 1)$ for all $k > k_n$ and all z with $|z| < \rho_n$. It then follows that (with the notation $\mathcal{D}_n := \{z;\ |z| < \rho_n\}$)

$$\forall k > k_n,\ \forall \pi \in \Pi(\mu,\nu): \quad \int_{\mathcal{D}\times\mathcal{D}} d^{\mathrm{H}}\big(m_k(z), w\big)\, d\pi(z,w)$$
$$\geqslant \int_{\mathcal{D}_n\times\mathcal{D}_n} d^{\mathrm{H}}\big(m_k(z), w\big)\, d\pi(z,w)$$
$$\geqslant \int_{\mathcal{D}_n\times\mathcal{D}_n} \Big[\inf_{|v|<\rho_n,\, |u|>(n+\rho_n)/(n+1)} d^{\mathrm{H}}(u,v)\Big]\, d\pi(z,w)$$
$$\geqslant \frac{\ln(n+1)}{2} \int_{\mathcal{D}_n\times\mathcal{D}_n} d\pi(z,w)$$
$$\geqslant \frac{\ln(n+1)}{2}\Big[1 - \pi\big((\mathcal{D}\setminus\mathcal{D}_n)\times\mathcal{D}\big) - \pi\big(\mathcal{D}\times(\mathcal{D}\setminus\mathcal{D}_n)\big)\Big]$$
$$\geqslant \frac{\ln(n+1)}{2}\left(1 - \frac{2}{n}\right),$$

where we have used that if $|u| < r < 1$, and $|v| > (n+r)/(n+1)$, then $d^{\mathrm{H}}(u,v) \geqslant \int_r^{(n+r)/(1+n)} \frac{1}{1-t^2}\, dt \geqslant \frac{1}{2}\int_r^{(n+r)/(1+n)} \frac{1}{1-t}\, dt = \frac{\ln(n+1)}{2}$. This shows, in particular, that

$$\mathrm{d}\big(\mu, m_k^*\nu\big) = \inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathcal{D}\times\mathcal{D}} d^{\mathrm{H}}\big(m_k(z), w\big)\, d\pi(z,w) \geqslant \frac{\ln(n+1)}{4}$$

for all $k > k_n$ and $n > 4$. This implies that, for arbitrary $n > 4$, $n \in \mathbb{N}$,

$$\mathcal{D}(\mu,\nu) = \lim_{k\to\infty} \mathrm{d}\big(\mu, m_k^*\nu\big) \geqslant \frac{\ln(n+1)}{4},$$

i.e. $\mathcal{D}(\mu,\nu) = \infty$, a contradiction. This finishes the argument that $|\bar{a}| = 1$ is not possible, and completes the proof. □

It is now easy to see that $\mathcal{D}(\mu,\nu)$ defines a true metric on the equivalence classes of conformal factors:

Proposition A.4. *The $\mathcal{D}(\mu,\nu)$ defined in* (3.1) *is a metric on the set of orbits $\overline{\mu} := \{m_*\mu;\ m \in M_{\mathcal{D}}\}$ of conformal factors under the action of $M_{\mathcal{D}}$.*

Proof. In view of Lemma A.1, we need to prove only that if $\mathcal{D}(\mu,\nu) = 0$ then there exists a Möbius transformation $m \in M_{\mathcal{D}}$ such that $\nu = m_*\mu$. By Lemma A.3, we know that $\mathcal{D}(\mu,\nu) = \mathrm{d}(m_*\mu,\nu)$ for some $m \in M_{\mathcal{D}}$, so that $\mathcal{D}(\mu,\nu) = 0$ implies $\mathrm{d}(m_*\mu,\nu) = 0$ for this m. Because d is a metric [14], it follows that $m_*\mu = \nu$. □

Next we prove the list of properties of the distance function $d^R_{\mu,\nu}(z,w)$ given in Theorem 3.3:

Theorem 3.3. *The distance function $d^R_{\mu,\nu}(z,w)$ satisfies the following properties*:

(1) $d^R_{m_1^*\mu,m_2^*\nu}\big(m_1^{-1}(z_0),m_2^{-1}(w_0)\big)=d^R_{\mu,\nu}(z_0,w_0)$ *invariance under* (*well-defined*) *Möbius changes of coordinates*;

(2) $d^R_{\mu,\nu}(z_0,w_0)=d^R_{\nu,\mu}(w_0,z_0)$ *symmetry*;

(3) $d^R_{\mu,\nu}(z_0,w_0)\geqslant 0$ *non-negativity*;

(4) $d^R_{\mu,\nu}(z_0,w_0)=0 \quad\Rightarrow\quad \Omega_{z_0,R}$ *in* $(\mathcal{D},\mu)$ *and* $\Omega_{w_0,R}$ *in* $(\mathcal{D},\nu)$ *are isometric*;

(5) $d^R_{m^*\nu,\nu}\big(m^{-1}(z_0),z_0\big)=0$ *reflexivity*;

(6) $d^R_{\mu_1,\mu_3}(z_1,z_3)\leqslant d^R_{\mu_1,\mu_2}(z_1,z_2)+d^R_{\mu_2,\mu_3}(z_2,z_3)$ *triangle inequality.*

Proof. For (1), denote $m_1^{-1}(z_0)=z_1$, and $m_2^{-1}(w_0)=w_1$. Then

$$\begin{aligned}d^R_{m_1^*\mu,m_2^*\nu}(z_1,w_1)&=\inf_{m(z_1)=w_1}\int_{\Omega_{z_1,R}}\big|m_1^*\mu(z)-m^*m_2^*\nu(z)\big|\,d_{\mathrm{vol}_H}(z)\\&=\inf_{m(z_1)=w_1}\int_{\Omega_{z_1,R}}\big|\mu\big(m_1(z)\big)-\nu\big(m_2\big(m(z)\big)\big)\big|\,d_{\mathrm{vol}_H}(z).\end{aligned}$$

Next set $\widetilde{m}=m_2\circ m\circ m_1^{-1}$. Note that $\widetilde{m}(z_0)=w_0$. Plugging $m_2(m(z))=\widetilde{m}(m_1(z))$ into the integral and carrying out the change of variables $m_1(z)=z'$, we obtain

$$\inf_{m(z_1)=w_1}\int_{\Omega_{z_0,R}}\big|\mu(z')-\nu\big(\widetilde{m}(z')\big)\big|\,d_{\mathrm{vol}_H}(z')=\inf_{\widetilde{m}(z_0)=w_0}\int_{\Omega_{z_0,R}}\big|\mu(z')-\nu\big(\widetilde{m}(z')\big)\big|\,d_{\mathrm{vol}_H}(z').$$

For (2), we use Lemma 3.2 and Eqs. (2.5), (2.6) to write

$$\begin{aligned}d^R_{\mu,\nu}(z_0,w_0)&=\inf_{m(z_0)=w_0}\int_{\Omega_{z_0},R}\big|\mu(z)-m^*\nu(z)\big|\,d_{\mathrm{vol}_H}(z)\\&=\inf_{m(z_0)=w_0}\int_{\Omega_{w_0},R}\big|\big(m^{-1}\big)^*\mu(w)-\nu(w)\big|\,d_{\mathrm{vol}_H}(w)\\&=d^R_{\nu,\mu}(w_0,z_0).\end{aligned}$$

(3) and (4) are immediate from the definition of $d^R_{\mu,\nu}$.

(5) follows from the observation that the minimizing m (in the definition (3.5) of $d^R_{\mu,\nu}$) is m_1 itself, for which the integrand, and thus the whole integral vanishes identically.

For (6), let m_1 be a Möbius transformation such that $m_1(z_1) = z_2$, and m_2 such that $m_2(z_2) = z_3$. Setting $m = m_2 \circ m_1$, we have

$$\begin{aligned} d^R_{\mu_1,\mu_3}(z_1, z_3) &\leqslant \int_{\Omega_{z_1,R}} \left|\mu_1(z) - m^*\mu_3(z)\right| d_{\mathrm{vol}_H}(z) \\ &\leqslant \int_{\Omega_{z_1,R}} \left|\mu_1(z) - m_1^*\mu_2(z)\right| d_{\mathrm{vol}_H}(z) \\ &\quad + \int_{\Omega_{z_1,R}} \left|m_1^*\mu_2(z) - m^*\mu_3(z)\right| d_{\mathrm{vol}_H}(z). \end{aligned} \tag{A.1}$$

The second term in (A.1) can be rewritten as (using Lemma 3.2, the change of coordinates $m_1(z) = w$ and the observation $m^* = m_1^* m_2^*$)

$$\begin{aligned} \int_{\Omega_{z_1,R}} \left|m_1^*\mu_2(z) - m^*\mu_3(z)\right| d_{\mathrm{vol}_H}(z) &= \int_{\Omega_{z_2,R}} \left|m_{1*}m_1^*\mu_2(w) - m_{1*}m_1^*m_2^*\mu_3(w)\right| d_{\mathrm{vol}_H}(w) \\ &= \int_{\Omega_{z_2,R}} \left|\mu_2(w) - m_2^*\mu_3(w)\right| d_{\mathrm{vol}_H}(w). \end{aligned}$$

We have thus

$$d^R_{\mu_1,\mu_3}(z_1, z_3) \leqslant \int_{\Omega_{z_1,R}} \left|\mu_1(z) - m_1^*\mu_2(z)\right| d_{\mathrm{vol}_H}(z) + \int_{\Omega_{z_2,R}} \left|\mu_2(w) - m_2^*\mu_3(w)\right| d_{\mathrm{vol}_H}(w),$$

and this for any $m_1, m_2 \in M_{\mathcal{D}}$ such that $m_1(z_1) = z_2$ and $m_2(z_2) = z_3$. Minimizing over m_1 and m_2 then leads to the desired result. □

Next we prove the continuity properties of the function $\Phi(z_0, w_0, \sigma) = \int_{\Omega(z_0,R)} |\mu(z) - \nu(m_{z_0,w_0,\sigma}(z))|\, d_{\mathrm{vol}_H}(z)$, stated in Lemma 3.6, which were used to prove continuity of $d^R_{\mu,\nu}$ itself (in Theorem 3.7).

Lemma 3.6.

- *For each fixed (z_0, w_0) the function $\Phi(z_0, w_0, \cdot)$ is continuous on S_1.*
- *For each fixed $\sigma \in S_1$, $\Phi(\cdot\,, \cdot, \sigma)$ is continuous on $\mathcal{D} \times \mathcal{D}$. Moreover, the family $(\Phi(\cdot\,, \cdot, \sigma))_{\sigma \in S_1}$ is equicontinuous.*

Proof. We start with the continuity in σ. We have

$$\left|\Phi(z_0, w_0, \sigma) - \Phi\big(z_0, w_0, \sigma'\big)\right| \leqslant \int_{\Omega(z_0,R)} \left|\nu\big(m_{z_0,w_0,\sigma}(z)\big) - \nu\big(m_{z_0,w_0,\sigma'}(z)\big)\right| d_{\mathrm{vol}_H}(z).$$

Because ν is continuous on $\mathcal{D}$, its restriction to the compact set $\overline{\Omega(w_0, R)}$ (the closure of $\Omega(w_0, R)$) is bounded. Since the hyperbolic volume of $\Omega(z_0, R)$ is finite, the integrand is dominated, uniformly in σ', by an integrable function. Since $m_{z_0,w_0,\sigma}(z)$ is obviously continuous in σ, we can use the dominated convergence theorem to conclude.

Since S^1 is compact, this continuity implies that the infimum in the definition of $d^R_{\mu,\nu}$ can be replaced by a minimum:

$$d^R_{\mu,\nu}(z_0, w_0) = \min_{m(z_0)=w_0} \int_{\Omega(z_0,R)} \big|\mu(z) - \nu\big(m(z)\big)\big|\, d\mathrm{vol}_H(z).$$

Next we prove continuity in z_0 and w_0 (with estimates that are uniform in σ).

Consider two pairs of points, (z_0, w_0) and $(z_0', w_0') \in \mathcal{D} \times \mathcal{D}$. Then

$$\begin{aligned}
&\big|\Phi(z_0, w_0, \sigma) - \Phi\big(z_0', w_0', \sigma\big)\big| \\
&\quad = \left| \int_{\Omega(z_0,R)} \big|\mu(z) - \nu\big(m_{z_0,w_0,\sigma}(z)\big)\big|\, d\mathrm{vol}_H(z) - \int_{\Omega(z_0',R)} \big|\mu(u) - \nu\big(m_{z_0',w_0'\sigma}(u)\big)\big|\, d\mathrm{vol}_H(u) \right| \\
&\quad \leqslant \left| \int_{\Omega(z_0,R)} \big|\mu(z) - \nu\big(m_{z_0,w_0,\sigma}(z)\big)\big| d\mathrm{vol}_H(z) \right. \\
&\qquad \left. - \int_{\Omega(z_0,R)} \big|\mu(m_{z_0,z_0',1}z) - \nu\big(m_{z_0',w_0',\sigma} \circ m_{z_0,z_0',1}(z)\big)\big|\, d\mathrm{vol}_H(z) \right| \\
&\quad \leqslant \int_{\Omega(z_0,R)} \Big(\big|\mu(z) - \mu\big(m_{z_0,z_0',1}(z)\big)\big| \\
&\qquad + \big|\nu\big(m_{z_0,w_0,\sigma}(z)\big) - \nu\big(m_{z_0',w_0',\sigma}\big(m_{z_0,z_0',1}(z)\big)\big)\big|\Big)\, d\mathrm{vol}_H(z).
\end{aligned}$$

On the other hand, note that for any $\gamma > 0$, μ and ν are continuous on the closures of $\Omega(z_0, R+\gamma)$ and $\Omega(w_0, R+\gamma)$, respectively; since these closed hyperbolic disks are compact, μ and ν are bounded on these sets. Pick now $\rho > 0$ such that $|z_0' - z_0| < \rho$, $|w_0' - w_0| < \rho$ imply that $\Omega(z_0', R) \subset \Omega(z_0, R+\gamma)$ as well as $\Omega(w_0', R) \subset \Omega(w_0, R+\gamma)$. It follows that, if $|z_0' - z_0| < \rho$ and $|w_0' - w_0| < \rho$, then $|\mu(z) - \mu(m_{z_0,z_0',1}(z))|$ and $|\nu(m_{z_0,w_0,\sigma}(z)) - \nu(m_{z_0',w_0',\sigma}(m_{z_0,z_0',1}(z)))|$ are bounded uniformly for $z \in \Omega(z_0, R)$. Since it is clear from the explicit expressions (3.7) that $m_{z_0,z_0',1}(z) \to z$ and $m_{z_0',w_0',\sigma}(m_{z_0,z_0',1}(z)) \to m_{z_0,w_0,\sigma}(z)$ as $z_0' \to z_0$ and $w_0' \to w_0$, we can thus invoke the dominated convergence theorem again to prove continuity of $\Phi(\cdot, \cdot, \sigma)$.

To prove the equicontinuity, we first note that ν is uniformly continuous on $\Omega(w_0, R) \cup \Omega(w_0', R)$, since ν is continuous on the compact set $\overline{\Omega(w_0, R+\gamma)}$, which contains $\Omega(w_0, R) \cup \Omega(w_0', R)$ for all w_0' that satisfy $|w_0' - w_0| \leqslant \rho$. This means that, given any $\varepsilon > 0$, we can find $\delta > 0$ such that $|\nu(w) - \nu(w')| \leqslant \varepsilon$ holds for all w, w' that satisfy $w, w' \in \Omega(w_0, R) \cup \Omega(w_0', R)$ and $|w - w'| \leqslant \delta$. This implies the desired equicontinuity if we can show that $|m_{z_0,w_0,\sigma}(z) - m_{z_0',w_0',\sigma}(m_{z_0,z_0',1}(z))|$ can be made smaller than δ, uniformly in $\sigma \in S_1$, by making $|z_0' - z_0| + |w_0' - w_0|$ sufficiently small.

We first estimate $|m_{z_0,w_0,\sigma}(z) - m_{z_0,w_0',\sigma}(z)|$. With the notations of (3.7), we have

$$a(z_0,w_0,\sigma) - a\big(z_0,w_0',\sigma\big) = \frac{(z_0 - w_0\overline{\sigma})(1-\overline{z_0}w_0'\overline{\sigma}) - (z_0 - w_0'\overline{\sigma})(1-\overline{z_0}w_0\overline{\sigma})}{(1-\overline{z_0}w_0\overline{\sigma})(1-\overline{z_0}w_0'\overline{\sigma})}$$
$$= \frac{(w_0 - w_0')\overline{\sigma}(|z_0|^2-1)}{(1-\overline{z_0}w_0\overline{\sigma})(1-\overline{z_0}w_0'\overline{\sigma})},$$

so that

$$\big|a(z_0,w_0,\sigma) - a\big(z_0,w_0',\sigma\big)\big| \leqslant \frac{|w_0-w_0'|}{(1-|z_0||w_0|)[1-|z_0|(|w_0|+\xi)]}$$
$$\leqslant \frac{\xi}{(1-|z_0||w_0|)[1-|z_0|(|w_0|+\xi)]}$$

when $|w_0 - w_0'| < \xi$. It thus suffices to choose ξ so that $\xi < \zeta(1-|z_0||w_0|)[1-|z_0|(|w_0|+\xi)]$ to ensure that $|a(z_0,w_0,\sigma) - a(z_0,w_0',\sigma)| < \zeta$. For the phase factor τ in (3.7) we obtain

$$\tau(z_0,w_0,\sigma) - \tau\big(z_0,w_0',\sigma\big)$$
$$= \sigma\frac{(1-\overline{z_0}w_0'\overline{\sigma})(1-z_0\overline{w_0}\sigma) - (1-\overline{z_0}w_0\overline{\sigma})(1-z_0\overline{w_0'}\sigma)}{(1-\overline{z_0}w_0\overline{\sigma})(1-\overline{z_0}w_0'\overline{\sigma})}$$
$$= \sigma\frac{(w_0-w_0')\overline{z_0}\overline{\sigma} - (\overline{w_0}-\overline{w_0'})z_0\sigma + |z_0|^2(\overline{w_0}w_0' - \overline{w_0'}w_0)}{(1-\overline{z_0}w_0\overline{\sigma})(1-\overline{z_0}w_0'\overline{\sigma})}$$
$$= \sigma\frac{(w_0-w_0')\overline{z_0}\overline{\sigma} - z_0(\overline{w_0}-\overline{w_0'})\sigma + |z_0|^2[\overline{w_0}(w_0'-w_0) + w_0(\overline{w_0}-\overline{w_0'})]}{(1-\overline{z_0}w_0\overline{\sigma})(1-\overline{z_0}w_0'\overline{\sigma})};$$

when $|w_0 - w_0'| < \xi$, this implies

$$\big|\tau(z_0,w_0,\sigma) - \tau\big(z_0,w_0',\sigma\big)\big| \leqslant \frac{|z_0||w_0|[2+|z_0|(2|w_0|+\xi)]}{(1-|z_0||w_0|)[1-|z_0|(|w_0|+\xi)]}\xi,$$

which can clearly be made smaller than any $\zeta > 0$ by choosing ξ sufficiently small. All this implies that (use (3.7))

$$\big|m_{z_0,w_0,\sigma}(z) - m_{z_0,w_0',\sigma}(z)\big|$$
$$\leqslant \big|\tau(z_0,w_0,\sigma) - \tau\big(z_0,w_0',\sigma\big)\big|\frac{1+|z|}{1-|z|} + \big|a(z_0,w_0,\sigma) - a\big(z_0,w_0',\sigma\big)\big|\frac{(1+|z|)^2}{(1-|z|)^2}$$
$$\leqslant \zeta\frac{2(1+|z|)}{(1-|z|)^2},$$

which will be smaller than $\delta/2$, uniformly in σ, if $\zeta < \delta(1-|z|^2)/8$; this bound on ζ in turn determines the bound to be imposed on the ξ used above. Hence $|m_{z_0,w_0,\sigma}(z) - m_{z_0,w_0',\sigma}(z)| < \delta/2$ can be guaranteed, uniformly in σ, by choosing $|w_0 - w_0'| < \xi$ for sufficiently small ξ.

One can estimate likewise

$$\left|m_{z_0,w_0',\sigma}(z) - m_{z_0',w_0',\sigma}\left(m_{z_0',z_0,1}(z)\right)\right|,$$

and show that this too can be made smaller than $\delta/2$, uniformly in σ, by imposing sufficiently tight bounds on $|z_0' - z_0|$ and $|w_0' - w_0|$. Combining all these estimates then leads to the desired equicontinuity, as indicated earlier. □

To prove Lemma 3.8, we shall use the following lemma:

Lemma A.5. *Consider $u_k = e^{\mathbf{i}\psi} + \varepsilon_k$, where $|\varepsilon_k| \to 0$ as $k \to \infty$. Then there exists, for every $\varepsilon > 0$, a $K \in \mathbb{N}$ such that for all $k > K$; and all $\widehat{m} \in M_{D,0,u_k}$,*

$$\inf_{w \in \Omega_{0,R}} \left|\widehat{m}(w)\right| > 1 - \varepsilon.$$

The set $M_{D,0,u_k}$ used in this lemma is given by Definition 3.4.

Proof. From Lemma 3.5 we can write $\widehat{m}$ as

$$\widehat{m}(w) = e^{\mathbf{i}\theta} \frac{w + u_k e^{-\mathbf{i}\theta}}{1 + \overline{u_k} e^{\mathbf{i}\theta} w},$$

for some $\theta \in [0, 2\pi)$. Substituting $u_k = e^{\mathbf{i}\psi} + \varepsilon_k$ in this equation we get

$$\widehat{m}(w) = e^{\mathbf{i}\theta} \frac{w + (e^{\mathbf{i}\psi} + \varepsilon_k)e^{-\mathbf{i}\theta}}{1 + \overline{(e^{\mathbf{i}\psi} + \varepsilon_k)} e^{\mathbf{i}\theta} w} = e^{\mathbf{i}\psi} \frac{1 + w e^{\mathbf{i}(\theta - \psi)} + \varepsilon_k \varepsilon^{-\mathbf{i}\psi}}{1 + w e^{\mathbf{i}(\theta-\psi)} + \overline{\varepsilon_k} e^{\mathbf{i}\theta} w}.$$

Writing the shorthand s for $s = 1 + w e^{\mathbf{i}(\theta-\psi)}$, we have thus

$$\begin{aligned}\left|\widehat{m}(w) - e^{\mathbf{i}\psi}\right| &= \left| e^{\mathbf{i}\psi} \frac{s + \varepsilon_k \varepsilon^{-\mathbf{i}\psi}}{s + \overline{\varepsilon_k} e^{\mathbf{i}\theta} w} - e^{\mathbf{i}\psi} \right| \leqslant \left| e^{\mathbf{i}\psi} \frac{\varepsilon_k e^{-\mathbf{i}\psi} - \bar{\varepsilon}_k e^{\mathbf{i}\theta} w}{s + \bar{\varepsilon}_k e^{\mathbf{i}\theta} w} \right| \\ &\leqslant \frac{|\varepsilon_k e^{-\mathbf{i}\psi} - \bar{\varepsilon}_k e^{\mathbf{i}\theta} w|}{|s + \bar{\varepsilon}_k e^{\mathbf{i}\theta} w|} \leqslant \frac{|\varepsilon_k|(1 + |w|)}{|s| - |\varepsilon_k||w|}\end{aligned}$$

Now for all $w \in \Omega_{0,R}$, $|w| < r_R = \tanh^{-1}(R)$. This implies $|s| \geqslant 1 - |w| \geqslant 1 - r_R$, and $1 + |w| \leqslant 1 + r_R$, so that

$$\left|\widehat{m}(w) - e^{\mathbf{i}\psi}\right| \leqslant |\varepsilon_k| \frac{1 + r_R}{1 - r_R - |\varepsilon_k| r_R} = |\varepsilon_k| \frac{1 + r_R}{1 - r_R(1 + |\varepsilon_k|)}.$$

Since $|\varepsilon_k| \to 0$ the lemma follows. □

We are now ready for

Lemma 3.8. *Let $\{(z_k, w_k)\}_{k \geqslant 1} \subset \mathcal{D} \times \mathcal{D}$ be a sequence that converges, in the Euclidean norm, to some point $(z', w') \in \overline{\mathcal{D}} \times \overline{\mathcal{D}} \setminus \mathcal{D} \times \mathcal{D}$, that is $|z_k - z'| + |w_k - w'| \to 0$, as $k \to \infty$. Then, $\lim_{k\to\infty} d^R_{\xi,\zeta}(z_k, w_k)$ exists and depends only on the limit point (z', w').*

Proof. Since $(z', w') \in \overline{\mathcal{D}} \times \overline{\mathcal{D}} \setminus \mathcal{D} \times \mathcal{D}$ either $z' \in \overline{\mathcal{D}} \setminus \mathcal{D}$ or $w' \in \overline{\mathcal{D}} \setminus \mathcal{D}$. Let us assume that $z' \in \overline{\mathcal{D}} \setminus \mathcal{D}$ (the case $w' \in \overline{\mathcal{D}} \setminus \mathcal{D}$ is similar). Denote by m_k an arbitrary Möbius transformation in $M_{D,0,w_k}$. By symmetry of the distance and using a change of variables we then obtain

$$\begin{aligned} d^R_{\xi,\zeta}(z_k, w_k) &= d^R_{\zeta,\xi}(w_k, z_k) \\ &= \min_{m(w_k)=z_k} \int\limits_{\Omega_{w_k,R}} \left|\zeta(w) - \xi\big(m(w)\big)\right| d_{\mathrm{vol}_H}(w) \\ &= \min_{m(w_k)=z_k} \int\limits_{\Omega_{0,R}} \left|\zeta\big(m_k(w)\big) - \xi\big(m\big(m_k(w)\big)\big)\right| d_{\mathrm{vol}_H}(w). \end{aligned}$$

Now, recall that $\xi(z) = \xi^H(z) = \widetilde{\xi}(z)(1 - |z|^2)^2$, where $\widetilde{\xi}(z)$ is a bounded function, $\sup_{z\in\mathcal{D}} |\widetilde{\xi}(z)| \leqslant C_{\widetilde{\xi}}$. From Lemma A.5 we know that for every $\varepsilon > 0$ and for $k > K$ sufficiently large, $|m(m_k((w)))| > 1 - \varepsilon$ for all $w \in \Omega_{0,R}$, and all m such that $m(w_k) = z_k$. This means that for these $k > K$ we have

$$\begin{aligned} \left|\xi\big(m\big(m_k(w)\big)\big)\right| &= \left|\widetilde{\xi}\big(m\big(m_k(w)\big)\big)\right|\big(1 - \left|m\big(m_k(w)\big)\right|^2\big)^2 \\ &\leqslant C_{\widetilde{\xi}}\big(1 - (1-\varepsilon)^2\big)^2 \leqslant C_{\widetilde{\xi}}\varepsilon^2(2-\varepsilon)^2, \end{aligned}$$

for all $w \in \Omega_{0,R}$. Therefore,

$$\begin{aligned} &\left| d^R_{\xi,\zeta}(z_k, w_k) - \int\limits_{\Omega_{0,R}} \left|\zeta\big(m_k(w)\big)\right| d_{\mathrm{vol}_H}(w) \right| \\ &\quad\leqslant \left| \min_{m(w_k)=z_k} \int\limits_{\Omega_{0,R}} \left|\zeta\big(m_k(w)\big) - \xi\big(m\big(m_k(w)\big)\big)\right| d_{\mathrm{vol}_H}(w) - \int\limits_{\Omega_{0,R}} \left|\zeta\big(m_k(w)\big)\right| d_{\mathrm{vol}_H}(w) \right| \\ &\quad\leqslant \left| \min_{m(w_k)=z_k} \int\limits_{\Omega_{0,R}} \left\{\left|\zeta\big(m_k(w)\big) - \xi\big(m\big(m_k(w)\big)\big)\right| - \left|\zeta\big(m_k(w)\big)\right|\right\} d_{\mathrm{vol}_H}(w) \right| \\ &\quad\leqslant \min_{m(w_k)=z_k} \int\limits_{\Omega_{0,R}} \left|\xi\big(m\big(m_k(w)\big)\big)\right| d_{\mathrm{vol}_H}(w) \to 0, \quad \text{as } k \to \infty. \end{aligned}$$

Therefore $d^R_{\xi,\zeta}(z_k, w_k)$ converges, as $k \to \infty$, if and only if $\int_{\Omega_{0,R}} |\zeta(m_k(w))|\, d_{\mathrm{vol}_H}(w)$ converges, and to the same limit, for any $m_k \in M_{D,0,w_k}$. We can take, for instance, $m_k(w) = \frac{w+w_k}{1+\overline{w_k}w}$ which gives

$$\int\limits_{\Omega_{0,R}} \left|\zeta\big(m_k(w)\big)\right| d_{\mathrm{vol}_H}(w) = \int\limits_{\Omega_{0,R}} \left|\zeta\left(\frac{w+w_k}{1+\overline{w_k}w}\right)\right| d_{\mathrm{vol}_H}(w).$$

For $w \in \Omega_{0,R}$, $|1+\overline{w_k}w| > 1-r_R$. It follows that this expression has a limit as $k \to \infty$, and

$$\lim_{k\to\infty} \int_{\Omega_{0,R}} \left|\zeta\left(m_k(w)\right)\right| d_{\mathrm{vol}_H}(w) = \int_{\Omega_{0,R}} \left|\zeta\left(\frac{w+w'}{1+\overline{w'}w}\right)\right| d_{\mathrm{vol}_H}(w),$$

which clearly depends on w', not on the sequence $\{w_k\}$. □

References

[1] D.M. Boyer, Y. Lipman, E.St. Clair, J. Puente, B.A. Patel, T.A. Funkhouser, J. Jernvall, I. Daubechies, New algorithms to automatically quantify the geometric similarity of anatomical surfaces, submitted for publication.

[2] E. Cela, The Quadratic Assignment Problem: Theory and Algorithms (Combinatorial Optimization), Springer, 1998.

[3] Facundo Memoli, On the use of Gromov–Hausdorff distances for shape comparison, in: Symposium on Point Based Graphics, 2007.

[4] Xianfeng David Gu, Shing-Tung Yau, Computational Conformal Geometry, International Press, Boston, 2008, Har/Cdr edition.

[5] L. Kantorovich, On the translocation of masses, C. R. Acad. Sci. URSS (N.S.) 37 (1942) 199–201.

[6] R. Kimmel, A.M. Bronstein, M.M. Bronstein, Generalized multidimensional scaling: a framework for isometry-invariant partial surface matching, Proc. Natl. Acad. Sci. 103 (5) (2006) 1168–1172.

[7] Irwin Kra, Hershel M. Farkas, Riemann Surfaces, Springer, 1992.

[8] Yaron Lipman, Ingrid Daubechies, Surface comparison with mass transportation, technical report, arXiv:0912.3488v2, 2009.

[9] Yaron Lipman, Thomas Funkhouser, Mobius voting for surface correspondence, in: Proc. SIGGRAPH, ACM Trans. Graph. 28 (3) (2009).

[10] Facundo Mémoli, Guillermo Sapiro, A theoretical and computational framework for isometry invariant recognition of point cloud data, Found. Comput. Math. 5 (3) (2005) 313–347.

[11] G. Monge, Mémoire sur la théorie des déblais et des remblais, in: Histoire de l'Académie Royale des Sciences de Paris, avec les Mémoires de Mathématique et de Physique pour la même année, 1781, pp. 666–704.

[12] Y. Rubner, C. Tomasi, L.J. Guibas, The earth mover's distance as a metric for image retrieval, Int. J. Comput. Vis. 40 (2) (2000) 99–121.

[13] George Springer, Introduction to Riemann Surfaces, AMS Chelsea Publishing, 1981.

[14] Cedric Villani, Topics in Optimal Transportation, Grad. Stud. Math., vol. 58, American Mathematical Society, March 2003.

[15] W. Zeng, X. Yin, Y. Zeng, Y. Lai, X. Gu, D. Samaras, 3d face matching and registration based on hyperbolic Ricci flow, in: CVPR Workshop on 3D Face Processing, 2008, pp. 1–8.

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