Computing the conformal barycenter

Jason Cantarella  Henrik Schumacher

February 28, 2022

Abstract

The conformal barycenter of a point cloud on the sphere at infinity of the Poincaré ball model of hyperbolic space is a hyperbolic analogue of the geometric median of a point cloud in Euclidean space. It was defined by Douady and Earle as part of a construction of a conformally natural way to extend homeomorphisms of the circle to homeomorphisms of the disk, and it plays a central role in Millson and Kapovich’s model of the configuration space of cyclic linkages with fixed edgelengths.

In this paper we consider the problem of computing the conformal barycenter. Abikoff and Ye have given an iterative algorithm for measures on $S^1$ which is guaranteed to converge. We analyze Riemannian versions of Newton’s method computed in the intrinsic geometry of the Poincaré ball model. We give Newton-Kantorovich (NK) conditions under which we show that Newton’s method with fixed step size is guaranteed to converge quadratically to the conformal barycenter for measures on any $S^d$ (including infinite-dimensional spheres). For measures given by $n$ atoms on a finite dimensional sphere which obey the NK conditions, we give an explicit linear bound on the computation time required to approximate the conformal barycenter to fixed error. We prove that our NK conditions hold for all but exponentially few $n$ atom measures. For all measures with a unique conformal barycenter we show that a regularized Newton’s method with line search will always converge (eventually superlinearly) to the conformal barycenter. Though we do not have hard time bounds for this algorithm, experiments show that it is extremely efficient in practice and in particular much faster than the Abikoff-Ye iteration.

MSC-2020 classification: 65D18, 65E10, 53-08

1 Introduction

In [1], Douady and Earle defined the conformal barycenter of a measure on $S^{d-1}$. Suppose we give the (open) ball $B^d$ the geometry of the Poincaré ball model of hyperbolic space. For each pair of points $w \in B^d$ and $x$ on the sphere $S^{d-1}$ at infinity, there is a unique geodesic joining $w$ to $x$. The unit tangent vector to this geodesic at $w$ will be called the director $V_x(w)$. The conformal barycenter (see Figure 1) of $\mu$ is the point $w_*$ where the weighted average of directors vanishes:

$$F_\mu(w_*) := \int_{S^{d-1}} V_x(w_*) \, d\mu(x) = 0.$$
The conformal barycenter is comparable to the geometric median [2] of a weighted point cloud \( x_1, \ldots, x_n \) in Euclidean space with weights \( \omega_1, \ldots, \omega_n \), which is (generically) the point \( w_* \) where \( \sum_{i=1}^n \omega_i (x_i - w_*) / |x_i - w_*| = 0 \). The conformal barycenter is clearly isometry-equivariant in hyperbolic geometry; this means that if \( w_* (\mu) \) is the conformal barycenter of \( \mu \) and \( \varphi \) is a Möbius transformation of \( \mathbb{B}^d \), then \( w_* (\varphi \# \mu) = \varphi (w_* (\mu)) \), where \( \varphi \# \mu \) is the push-forward of \( \mu \). Douady and Earle’s original motivation for their construction was to provide a canonical extension of homeomorphisms of the circle to homeomorphisms of the disk that is also conformally equivariant.

The conformal barycenter for discrete measures composed of \( n \) atoms with weights \( \omega_1, \ldots, \omega_n \) plays a central role in the symplectic model for the space of closed polygons with fixed edgelengths in \( \mathbb{R}^3 \) given by Kapovich and Millson [3]. They identify the atoms \( x_1, \ldots, x_n \) of \( \mu \) with the directions of the edges of the polygon and \( \omega_1, \ldots, \omega_n \) with the lengths of the edges. Such a polygon is closed if and only if \( \sum_i \omega_i x_i = 0 \). They show that for a “stable” (see Definition 5.1) measure \( \mu \) of this type, there is a Möbius transformation \( \varphi \) so that the conformal barycenter \( w_* (\varphi \# \mu) = 0 \). Further, this \( \varphi \) is unique up to postcomposition with an element of \( \text{SO}(3) \), and \( w_* (\mu) = 0 \) if and only if \( \sum_i \omega_i x_i = 0 \).

This construction allows them to show (Theorem 2.7) that the quotient space of “nice semi-stable” (Definition 5.1) \( n \)-atom measures \( \mu \) with respect to all Möbius transformations is homeomorphic (and even complex-analytically equivalent) to the quotient space of \( n \)-edge closed polygons with edgelengths \( \omega_1, \ldots, \omega_n \) by the action of \( \text{SO}(3) \).

Generalizing their construction in the obvious way, this means that the space of nice semi-stable measures \( \mu \) (or weighted point clouds on \( \mathbb{S}^{d-1} \)) is an exact (but redundant) system of coordinates for the space of closed polygonal linkages with fixed edgelengths in \( \mathbb{R}^d \). That is, every stable weighted point cloud \( \mu \) on \( \mathbb{S}^{d-1} \) exactly represents a unique (up to the action of \( \text{SO}(d) \)) closed polygon in \( \mathbb{R}^d \) with the corresponding edgelengths. (We lose complex-analytic equivalence, as the spaces do not generally carry a complex structure.)

Such a system of coordinates promises to be extremely useful in studying linkage problems.
Figure 2: Millson and Kapovich’s construction: (a) An open polygon. (b) The corresponding point measure \( \mu \) on \( S^1 \) where the masses \( x_i \) of the measure are located at the directions of the edges (and weighted by their lengths), together with the conformal barycenter \( w_* \) of \( \mu \). (c) We see \( \mu \) shifted by a conformal transformation \( \varphi \) to bring the conformal barycenter (and hence the center of mass) to the origin. (d) The transformed edges make up a closed polygon with the same edgelengths.

in computational geometry and in implementing robust numerical methods for linkage reconfiguration. The linkage reconfiguration problem has a vast array of applications, including robotics, protein folding, polymer science and computational origami (see [4] for an overview of the field). However, to make these coordinates computationally practical, we must solve one remaining problem: If the \( \mu \) representing our polygon has \( \sum \omega_i x_i \neq 0 \), to recover the unique closed polygon represented by \( \mu \) we must find the Möbius transformation \( \varphi \) so that \( w_*(\varphi(\mu)) = 0 \). Since \( w_*(\varphi_\#\mu) = \varphi(w_*(\mu)) \), this is equivalent to finding \( w_*(\mu) \). Figure 2 shows an example of a stable weighted point cloud \( \mu \) on \( S^1 \) with \( \sum \omega_i x_i \neq 0 \), the corresponding open polygon, the Möbius transformed \( \varphi_\#\mu \) with \( \sum \omega_i \varphi(x_i) = 0 \), and the corresponding closed polygon.

This motivates us to consider the question of numerically computing the conformal barycenter of a measure on \( S^{d-1} \) and to pay particular attention to the case of \( n \)-atom measures with variable weights. Abikoff and Ye [5] and Abikoff [6] have given an iterative algorithm based on a suggestion of Milnor and an algorithm based on Newton’s method in the Euclidean geometry of \( \mathbb{R}^2 \) for computing the conformal barycenter of such a measure. However, they only prove that their method converges for measures on \( S^1 \) and were not able to give effective bounds for the rate of convergence. In practice, their method can be very slow for highly concentrated measures. The primary insight of this paper is that switching the setting of Newton’s method from the Euclidean geometry of the unit ball to the hyperbolic geometry of the Poincaré ball model leads to greatly improved theoretical and numerical results.

After recording some preliminary calculations (Section 2), we will analyze the Riemannian Newton’s method with fixed step size (Section 3). We give Newton-Kantorovich (NK) conditions under which the iteration converges quadratically for all steps (Theorem 3.2) using a theorem of Ferreira and Svaiter [7]. Our convergence results hold for all \( d \), including infinite-dimensional spheres. We then show that under NK conditions, our algorithm solves the problem “compute the conformal barycenter of an \( n \)-atom measure on \( S^d \) to within error \( \varepsilon \)” in \( O(n) \) time (Corollary 3.3).

We then turn to the question of how often Corollary 3.3 applies to \( n \) atom measures (Section 4). The answer is “in all but exponentially few (in \( n \)) cases” (Theorem 4.1). The key tool is an
estimate on the eigenvalues of a certain random matrix previously computed by one of us (Cantarella) using the Matrix Bernstein inequality [8].

We then consider cases where the NK conditions may not hold (Section 5). We analyze the behavior of a regularized Newton’s method with line search (29)–(30). We show that if \( \mu \) has a unique conformal barycenter \( w_\ast \), the algorithm converges (eventually Q-quadratically) to \( w_\ast \) (Theorem 5.9) using results of Ring and Wirth on Riemannian Newton methods [9]. The key idea is to recast the problem as finding the minimizer of a function \( \Psi_\mu \) (25) which we can prove to be uniformly convex on balls of finite radius in hyperbolic geometry (Lemma 5.6). Again, our results prove convergence and bound the rate for all finite \( d \) and even for infinite dimensional spheres.

We conclude by using our algorithm to compute examples of polygon closures and Douady-Earle extensions of maps from \( S^1 \) into \( S^2 \) (Section 6). Our algorithm performs very well even when the Milnor-Abikoff-Ye iteration struggles to converge.

2 Background

Although we currently only have applications for the computation of conformal barycenters for measures on \( S^{d-1} \) where \( d \) is finite, all our methods work just as well when the ambient space is an arbitrary Hilbert space. Therefore, we will work in that context for generality.

Let \( H \) be a real Hilbert space with inner product \( \langle \cdot, \cdot \rangle : H \times H \rightarrow \mathbb{R} \), norm \( |u| := \sqrt{\langle u, u \rangle} \), and with Riesz isomorphism \( (\cdot)^\top : H \rightarrow H' \) given by \( u^\top(v) := \langle u, v \rangle \) for \( u, v \in H \). We now recall the Poincaré ball model for hyperbolic space on \( H \). \( B \) will denote the (open) ball \( \{ z \in H \mid |z| < 1 \} \) with Riemannian metric \( g : B \rightarrow L^2(TB; \mathbb{R}) \) given by

\[
g_w(X, Y) := \frac{4}{(1-|w|^2)^2} \langle X, Y \rangle, \quad \text{for } w \in B \text{ and } X, Y \in T_w B \cong H. \tag{1}\]

The length of a tangent vector \( X \in T_w B \) with respect to \( g \) will be denoted by \( |X|_g := \sqrt{g_w(X, X)} \); we have to distinguish it carefully from \( |X| \). The induced Riesz isomorphism will be denoted by \( (\cdot)^\flat : T_w B \rightarrow T_w' B, X \mapsto X^\flat \) and the geodesic distance between two points \( w_1, w_2 \in B \) will be denoted by \( d_g(w_1, w_2) \). We point out that even if \( H \) is infinite-dimensional, the Poincaré ball model \( (B, g) \) is a (strongly) Riemannian manifold in the sense that \( g_w \) turns \( T_w B \) into a Hilbert space (see, e.g., [10, Chapter VII]). Moreover, \( (B, g) \) is geodesically complete and geodesically convex, properties inherited basically from the two-dimensional Poincaré model. One can deduce these properties also from the observation that the Riemannian exponential map \( \exp_0 : T_0 B \rightarrow B \) is just a reparameterized linear ray of the form

\[
\exp_0(X) = \frac{\tanh(|X|_g/2)}{|X|_g/2} X = \frac{\tanh(|X|)}{|X|} X \quad \text{for } X \in T_0 B \cong H. \tag{2}\]

Thus, geodesics emanating from the origin are just reparameterized straight lines, and geodesic completeness and geodesic convexity follow directly from the transitivity of the isometry group \( \text{Aut}(B, g) \) (see (2.1)) and from the fact that straight lines can intersect at most once (due to negative curvature).

The unit sphere in \( H \) will be denoted by \( S := \{ w \in H \mid |w| = 1 \} \). Likewise, the unit sphere with respect to the Poincaré metric \( g \) in the tangent space \( T_w B \) will be denoted by \( \text{UT}_w B := \)
\( \{ X \in T_w \mathbb{B} \mid |X|^g = 1 \} \). For each \( w \in \mathbb{B} \) and each \( x \in \mathbb{S} \), there is a unique unit tangent vector \( V_x(w) = V(w, x) \in UT_w \mathbb{B} \) satisfying
\[
\lim_{t \to \infty} \exp_t(V_x(w)) = x,
\]
where \( \exp: T \mathbb{B} \to \mathbb{B} \) denotes the Riemannian exponential map with respect to the metric \( g \) and where the limit is to be interpreted as limit in the topology of \( H \). Soon, we will derive a concrete expression for \( V_x(w) \) and realize that the \textit{director mapping} \( V: \mathbb{B} \times \mathbb{S} \to UT \mathbb{B} \) is a smooth mapping into the sphere bundle \( UT \mathbb{B} \) (see Section 2.2). For now, we define:

\textbf{Definition 2.1} Let \( \mu \) be a Borel probability measure on \( \mathbb{S} \). We say \( w \in \mathbb{B} \) is a \textit{conformal barycenter} of \( \mu \) if and only if
\[
F_\mu(w) = 0, \quad \text{where} \quad F_\mu(w) := \int_\mathbb{S} V_x(w) \, d\mu(x).
\]
We call \( \mu \) \textit{conformally centralized} if \( F_\mu(0) = 0 \).

Our aim is to derive conditions on \( \mu \) that ensure that \( F_\mu(w) = 0 \) has a unique solution \( w_\ast \) (in which case we are allowed to talk about the \textit{conformal barycenter}) and use Newton’s method in hyperbolic space to compute \( w_\ast \) iteratively.

\subsection*{2.1 Shift Transformation}

From (2), it follows that \( V_x(0) = \frac{1}{2} x \). In order to obtain an explicit expression for \( V_x(w) \) for general \( w \in \mathbb{B} \), it is useful to understand the isometry group \( \text{Aut}(\mathbb{B}, g) \) of the Riemannian manifold \( (\mathbb{B}, g) \) first. This group is characterized as the group of those Möbius transformations of \( H \cup \{|\infty|\} \) that map the unit ball \( \mathbb{B} \) onto itself. This is precisely the group generated by inversions in those spheres that meet \( \mathbb{S} \) perpendicularly. Such spheres cannot have their center on \( \mathbb{S} \). Hence each isometry \( \varphi: \mathbb{B} \to \mathbb{B} \) induces also a unique conformal diffeomorphism \( \varphi: \mathbb{S} \to \mathbb{S} \). By the construction of \( V \), we have for each \( \varphi \in \text{Aut}(\mathbb{B}, g) \) that
\[
V_{\varphi(x)}(\varphi(w)) = d\varphi(w) V_x(w) \quad \text{for all} \quad w \in \mathbb{B} \quad \text{and} \quad x \in \mathbb{S}.
\]
This induces the following transformation rule for \( F \) under \( \varphi \in \text{Aut}(\mathbb{B}, g) \):
\[
F_{\varphi_\mu}(\varphi(w)) = d\varphi(w) F_\mu(w).
\]
If we find an isometry \( \varphi \) that maps a conformal barycenter \( w_\ast \) of \( \mu \) (if existent) to 0, we can construct a centralized measure \( \varphi_\# \mu \) as we have
\[
F_{\varphi_\mu}(0) = F_{\varphi_\mu}(\varphi(w_\ast)) = d\varphi(w_\ast) F_\mu(w_\ast) = 0.
\]
To this end, we associate a unique hyperbolic translation \( \sigma_\ast \in \text{Aut}(\mathbb{B}, g) \) with each \( w \in \mathbb{B} \). It is characterized by \( \sigma_\ast(w) = 0 \) and can be expressed by
\[
\sigma: \mathbb{B} \times \mathbb{B} \to \mathbb{B}, \quad \sigma(w, z) := \sigma_\ast(w, z) := \frac{(1 - |w|^2) z - (1 + |z|^2 - 2 \langle w, z \rangle) w}{1 - 2 \langle w, z \rangle + |w|^2 |z|^2}.
\]
In the special case \( H = \mathbb{R}^2 \cong \mathbb{C} \), we may rewrite \( \sigma \) in complex arithmetic as
\[
\sigma(w, z) = \frac{\bar{w} z - \bar{z} w}{1 - \bar{w} \bar{z}}.
\]
This reveals that \( \sigma \) is merely an orientation-preserving version of the map \( \eta \) from [6].

where the limit is taken with respect the topology of $\mathbb{B}$. We may assume that $H$ is an open set, we have a canonical identification $T_x\mathbb{B} \cong H$. If we let $C(w, z) = (1 - |w|^2)(1 - 2 \langle w, z \rangle + |w|^2|z|^2)^{-1}$, then we can easily check that $\sigma_w$ is orientation preserving as its differential is given by

$$
d\sigma_w(z) = C(w, z) \left( id_H - 2 \frac{|z|^2 w w^T + (w z^T - z w^T - 2 \langle w, z \rangle w z^T + |w|^2 z z^T)}{1 - 2 \langle w, z \rangle + |w|^2} \right). \quad (7)
$$

It is now straightforward to compute that $d\sigma_w(z)(d\sigma_w(z))^T = C^2(w, z) id_H$, and hence that $d\sigma_w(z)$ is a similarity matrix with conformal factor $C(w, z)$. The mapping $\sigma_w$ is a hyperbolic translation in the sense that it maps the geodesic though $w$ and 0 to itself; more precisely, we have $\sigma(w, tw) = (1 - t)(1 - t|w|^2)^{-1} w$. Its inverse has the simple form $(\sigma_w)^{-1} = \sigma_{-w}$. For each linear subspace $E \subset H$ and for $w \in E \cap \mathbb{B}$, the transformation $\sigma_w$ maps $E \cap \mathbb{B}$ to itself. As for all elements of $Aut(\mathbb{B}, g)$, $\sigma_w$ can be extended to $\mathbb{S}$, and we have

$$
\sigma_w(x) := \sigma(w, x) := \lim_{z \to x} \sigma(w, z) = \frac{(1 - |w|^2) x - 2(1 - \langle w, x \rangle) w}{1 - 2 \langle w, x \rangle + |w|^2}, \quad (8)
$$

where the limit is taken with respect the topology of $H$. A short calculation reveals that $\sigma_w$ indeed maps $\mathbb{S}$ to $\mathbb{S}$. There is a nice geometric interpretation of (8): if we extend the (Euclidean) line from $x$ through $w$ until it strikes $\mathbb{S}$ at some point $p$, then $\sigma_w(x) = -p$. We will leave the proof to the reader (see also Figure 1, (c)), but we now use this observation to establish:

**Lemma 2.2** For $w_1, w_2 \in \mathbb{B}$, $x \in \mathbb{S}$, we have $d_2(\sigma_{w_1}(x), \sigma_{w_2}(x)) \leq 2d_2(w_1, w_2)$.

**Proof.** We may focus our attention to the linear space spanned by $w_1$, $w_2$, and $x$, so without loss of generality, we may assume that $H$ is three-dimensional. Let $C_i$ denote the Euclidean secant through $x$ containing $w_i$ and let $p_i$ be the other end point of $C_i$. The plane containing $w_1$, $w_2$ and $x$ cuts $\mathbb{S}$ in a circle with center $z$ and radius $r \leq 1$. From our observation above, we know that $\sigma_w(x) = -p_i$, so the central angle theorem implies

$$
d_2(\sigma_{w_1}(x), \sigma_{w_2}(x)) = d_2(-p_1, -p_2) = d_2(p_1, p_2) \\
\leq r \angle(p_1, z, p_2) \leq \angle(p_1, z, p_2) = 2 \angle(p_1, x, p_2) = 2 \angle(w_1, x, w_2),
$$

where $r = 1$.
leaving us to show $\angle(w_1, x, w_2) \leq d_\theta(w_1, w_2)$.

It suffices to do so for the maximizer $x = x_\ast$ of $\theta(x) := \angle(w_1, x, w_2)$ on $\mathbb{S}$. Observe that $\theta$ is continuous on $\mathbb{S}$ (the only discontinuities of $\theta$ are $w_1$ and $w_2$). So by compactness, a global maximizer $x_\ast$ of $\theta$ on $\mathbb{S}$ must exist. Next we show that $w_1$, $w_2$, $x_\ast$, and $0$ are coplanar: Denote the (Euclidean) straight line through $w_1$ and $w_2$ by $L$. The minimizers of $\theta$ on $\mathbb{S}$ are precisely the two points in $L \cap \mathbb{S}$, and $\theta$ is smooth away from $L$. Hence $x_\ast$ must be a critical point of $\theta|_L$.

This means that the surface normals $v_\ast(x_\ast) = x_\ast$ of $\mathbb{S}$ and $v_2(x_\ast)$ of the levelset $\Sigma$ of $\theta$ at the point $x_\ast$ are colinear. Levelsets of $\theta$ are surfaces of revolution about $L$, and so $v_2(x_\ast)$ is in the plane spanned by $x_\ast$, $w_1$ and $w_2$. By collinearity, this means that $v_\ast(x_\ast) = x_\ast$ is also in this plane, which must then pass through the origin.

Hence it suffices to show $\theta(x) \leq d_\theta(w_1, w_2)$ in the case that $H$ is the Euclidean plane. Denote the diameter from $x$ to $-x$ by $G$. Then there are two cases:

**Case 1:** $G$ lies between $C_1$ and $C_2$. Denote the geodesic connecting $x$ and $p_i$ by $G_i$ and the angle at $x$ that is enclosed by $G_i$ and $C_i$ by $\alpha_i$ (see Figure 3 (a)). Each $C_i$ is a hypercycle or equidistant curve from the corresponding geodesic $G_i$. This means that each point $w \in C_i$ has the same minimal distance $\text{dist}_g(w_i, G_i) = f(\alpha_i)$ to $G_i$ (see [11, Chapter 14, Theorem C]). It is known that $f(\alpha) = \text{arsinh}(\tan \alpha)$ (see [12, p. 24]). Because $G$ lies between $C_1$ and $C_2$, $G_1$ and $G_2$ also lie between $C_1$ and $C_2$, and we have

$$d_\theta(w_1, w_2) \geq \text{dist}_g(w_1, G_1) + \text{dist}_g(w_2, G_2) = f(\alpha_1) + f(\alpha_2).$$

Since $f(0) = 0$ and $f'(\alpha) = \sec \alpha \geq 1$ we have $f(\alpha) \geq \alpha$ for all $\alpha \in [0, \pi/2)$. Thus in particular, we have $f(\alpha_1) + f(\alpha_2) \geq \alpha_1 + \alpha_2 = \theta(x)$.

**Case 2:** $G$ does not lie between $C_1$ and $C_2$. Without loss of generality, we may assume that $C_1$ lies between $G$ and $C_2$ (see Figure 3 (b)). Denote the angle at $x$ between $G$ and $C_1$ by $\omega$. Observe that $C_1$ lies also between $G_1$ and $C_2$. Let $C$ be the circle through $x$ and $p_1$ that is tangent to $C_2$ at $x$. Again, $C$ is equidistant from $G_1$, meeting $G_1$ at the angle $\omega + \theta(x)$. Moreover, $C$ is equidistant from $C_1$ as well (see again [11, Chapter 14, Theorem C]). Hence we have

$$d_\theta(w_1, w_2) \geq \text{dist}_g(w_1, G) = \text{dist}_g(C, G_1) - \text{dist}_g(C_1, G_1) = f(\omega + \theta(x)) - f(\omega).$$

Since $f' \geq 1$, by the mean value theorem $f(\omega + \theta(x)) - f(\omega) \geq (\omega + \theta(x)) - \omega = \theta(x)$.

**2.2 Director Field**

Utilizing the family of isometries $\sigma$ and the identity (4), we may compute $V_\ast(w)$ for arbitrary $w \in \mathbb{B}$ by first moving $w$ to 0 by the shift transformation $\sigma_w$ (and by moving $x$ along with $\sigma_w$), by evaluating $V$ there, and by transporting the result back to $w$ with $d_\sigma(\sigma_w(0))$:

$$V_\ast(w) = d_\sigma(\sigma_w(0)) : V(0, \sigma_w(x)) = \frac{1}{2} (1 - |w|^2) \sigma_w(x),$$

where we used (7). Combined with (6), we obtain the explicit expression

$$V_\ast(w) = \frac{1}{2} \frac{(1 - |w|^2)^2 x - 2 (1 - |w|^2)(1 - \langle w, x \rangle) w}{1 - 2 \langle w, x \rangle + |w|^2}.$$  

(9)

Because of $1 - 2\langle w, x \rangle + |w|^2 = |x - w|^2 > 0$ for all $w \in \mathbb{B}$ and $x \in \mathbb{S}$, the mapping $V : \mathbb{B} \times \mathbb{S} \to \text{UT} \mathbb{S}$ is smooth.
2.3 Approximating conformally centralized measures

We have already seen above that shifting $\mu$ by one of its conformal barycenters $w$, leads to the centered measure $(\sigma_w)_#\mu$. We now show that a good approximation $w$ of $w_*$ will also lead to a good approximation $(\sigma_w)_#\mu$ of $(\sigma_{w_*})_#\mu$. We recall a definition first.

**Definition 2.3** Given two probability measures $\mu_1$ and $\mu_2$ on a space $X$, a transport plan $\gamma$ between the $\mu_i$ is a probability measure $\gamma$ on $X \times X$ with marginals equal to the $\mu_i$. The Wasserstein distance $W_p$ is defined by

$$W_p(\mu_1, \mu_2) = \begin{cases} \inf_{\gamma} \left( \int_{X \times X} d_\delta(y, z)^p \, d\gamma(y, z) \right)^{1/p}, & \text{for } 1 \leq p < \infty, \\ \inf_{\gamma} \text{ess sup} \ d_\delta(y, z), & \text{for } p = \infty, \end{cases}$$

where in each case the infimum runs over all transport plans $\gamma$ between $\mu_1$ and $\mu_2$.

We can now estimate the Wasserstein distance between shifts of a measure $\mu$:

**Lemma 2.4** Let $\mu$ be a Borel probability measure on $\mathbb{S}$. Then for all $w_1$ and $w_2 \in \mathbb{B}$, we have the following estimate for the $p$-Wasserstein distance with respect to the angular distance function $d_\delta$ on $\mathbb{S}$:

$$W_p((\sigma_{w_1})_#\mu, (\sigma_{w_2})_#\mu) \leq 2 d_\delta(w_1, w_2) \quad \text{for all } p \in [1, \infty]. \quad (10)$$

**Proof.** We abbreviate $\psi_i := \sigma_{w_i}$ and $\mu_i := (\sigma_{w_i})_#\mu = (\sigma_{w_i})_#\mu$, and we denote the projection onto the $i$-th Cartesian factor of $\mathbb{S} \times \mathbb{S}$ by $\pi_i: \mathbb{S} \times \mathbb{S} \to \mathbb{S}$. If we let $\Phi(x) = (\psi_1(x), \psi_2(x))$ and write $\nu = \Phi_#\mu$, then $(\pi_i)_#\nu = (\pi_i)_#\Phi_#\mu = (\pi_i \circ \Phi)_#\mu = (\psi_i)_#\mu = \mu_i$. Thus $\nu$ is a transport plan between the $\mu_i$ and

$$W_p(\mu_1, \mu_2) \leq W_\infty(\mu_1, \mu_2) \leq \text{ess sup}_{(y, z) \in \text{supp}(\nu)} d_\delta(y, z) = \sup_{x \in \mathbb{S}} d_\delta(\psi_1(x), \psi_2(x)),$$

and the result then follows from Lemma 2.2. \qed

3 Newton’s Method with fixed step size

We are now going to use Newton’s method with fixed step size in the Riemannian manifold $\mathbb{B}$ with metric $g$ (c.f. [7, Algorithm 1.2]) to solve the equation (3).

3.1 Basic algorithm

Starting with an initial guess $w_0 \in \mathbb{B}$, we generate a sequence of $w_k \in \mathbb{B}$ by

$$v_k := -\nabla F_\mu(w_k)^{-1} F_\mu(w_k) \quad \text{and} \quad w_{k+1} := \exp_{w_k}(v_k). \quad (11)$$

In order to derive an expression for $\nabla F_\mu$, we first compute the covariant derivative $\nabla V_\mu$ with respect to the metric $g$. It helps to recall that if we have conformally equivalent Riemannian
metrics $g(X, Y) = f(w) \bar{g}(X, Y)$ with conformal factor $f(w) > 0$, then the corresponding Riemannian connections $\nabla$ and $\bar{\nabla}$ are related by
\[
\nabla_X Y = \bar{\nabla}_X Y + (2 f)^{-1}((X f) Y + (Y f) X - \bar{g}(X, Y) \text{grad} \ f),
\]
where $\text{grad} \ f$ obeys $X f = \bar{g}(X, \text{grad} \ g)$ (cf. [13], p. 181). By (1), our metric $g$ is conformally equivalent to the metric $\bar{g} = (\cdot, \cdot)$ on $\mathbb{B}$ with conformal factor $f(w) = 4 (1 - |w|^2)^{-2}$. Since the covariant derivative $\bar{\nabla}$ coincides with the Fréchet derivative $D$, this (after a short computation) yields
\[
\nabla_X Y(w) = D Y(w) X(w) + \frac{2}{1 - |w|^2} (\langle X, w \rangle Y + \langle Y, w \rangle X - \langle X, Y \rangle w).
\]
Notice that the two covariant derivatives coincide at $w = 0$. Differentiating (9), it is straightforward to compute
\[
\nabla V_x(0) = D V_x(0) = x x^T - \text{id}_H = V_x(0) \otimes V_x(0) - \text{id}_{T_x \mathbb{B}}.
\]
To compute $\nabla V_x(w)$, we can either pull back the computation to $w = 0$ using covariance or use (13) directly at $w$ and simplify. Either way, we obtain
\[
\nabla V_x = V_x \otimes V_x^0 - \text{id}_{T_x \mathbb{B}}.
\]
Differentiating (3) under the integral sign, we derive from (14) that
\[
\nabla F_{\mu} = \int_{x \in \mathbb{S}} (V_x \otimes V_x^0) \, d\mu(x) - \text{id}_{T_x \mathbb{B}}.
\]

### 3.2 Shifted algorithm

When performing (11) by computing (9), (14), and $\exp_{w_k}(v_k)$ in machine arithmetic, one is frequently confronted with catastrophic loss of precision. We can avoid this by using an idea of Abikoff and Ye [5]: shift everything at point $w_k$ to the Euclidean origin, perform the Newton update there, and shift everything back to $w_k$. This gives us:
\[
\mu_k := (\sigma_{w_k})_# \mu, \quad u_k := -\nabla F_{\mu_k}(0)^{-1} F_{\mu_k}(0), \quad \text{and} \quad w_{k+1} = \sigma(-w_k, \exp_0(u_k)).
\]
Indeed, with $\varphi_k := \sigma(-w_k)$ and by (5), the search directions $v_k$ and $u_k$ are related as follows:
\[
v_k = -(\nabla F_{\mu_k}^{-1} F_{\mu_k})|_{w_k} = -((\nabla F_{\varphi_k}^{-1} F_{\mu_k} \circ \varphi_k)(F_{\varphi_k}^{-1} F_{\mu_k} \circ \varphi_k))|_0
\]
\[
= -((d\varphi_k \nabla F_{\mu_k} d\varphi_k^{-1})^{-1} d\varphi_k F_{\mu_k})|_0 = -d\varphi_k F_{\mu_k}^{-1} F_{\mu_k} |_0 = d\varphi_k(0) u_k.
\]

Together with $\exp_{w_k}(v_k) = \exp_{\varphi_k(0)}(d\varphi_k u_k) = \varphi_k(\exp_0(u_k))$, this shows that both (11) and (16) produce the same new iterate $w_{k+1}$—at least in exact arithmetic.

Computing $F_{\mu_k}(0)$ and $\nabla F_{\mu_k}(0)$ is particularly easy as we have
\[
F_{\mu_k}(0) = \int_0^1 s \, d\mu_k(x) \quad \text{and} \quad \nabla F_{\mu_k}(0) = D F_{\mu_k}(0) = \int_0^1 (x x^T) \, d\mu_k(x) - \text{id}_H.
\]

The exponential map $\exp_0 : T_0 \mathbb{B} \to \mathbb{B}$ can be computed according to (2) by
\[
\exp_0(u_k) = \frac{\tanh(|u_k|)}{|u_k|} u_k.
\]

Up to this point, the measure $\mu$ and the Hilbert space $H$ have been arbitrary. We could have given $\mu$ by any density supported on $\mathbb{S}$, computed $F_{\mu_k}(0)$ and $\nabla F_{\mu_k}(0)$ by numerical integration, and computed $\mu_{k+1}$ by the transformation formula for measures. However, for finite, discrete measures, the pushforward is much easier to compute as we see in the following example:
Example 3.1 Let $H = \mathbb{R}^d$ be a finite-dimensional Euclidean space and let \( \mu \) be given as a linear combination of Dirac measures $\mu = \sum_{i=1}^{n} \omega_i \delta(x_i)$ with $x_i \in \mathbb{R}^{d-1}$, $\omega_i \geq 0$, and $\sum_{i=1}^{n} \omega_i = 1$. Then, noting that at $w = 0$, we have $V_\ast(w) = \frac{1}{2} x$ and $d'(v) = g(u, v) = 4 \langle u, v \rangle_{\mathbb{R}^d} = 4 u^T v$, we get 
\[
F_\mu(0) = \frac{1}{2} \sum_{i=1}^{n} x_i \omega_i \quad \text{and} \quad \nabla F_\mu(0) = (\sum_{i=1}^{n} (x_i x_i^T) \omega_i) - \text{id}_{\mathbb{R}^d}.
\]
The push-forward $(\sigma_\ast)_\# \mu$ of the measure $\mu$ has the same weights $\omega_i$, but the locations of its Dirac measures are transformed by $\sigma_\ast$. It can be written as 
\[
(\sigma_\ast)_\# \mu = \sum_{i=1}^{n} \omega_i \delta(\sigma(s, x_i)).
\]

3.3 Statement of first convergence theorem

We will see now that Ferreira and Svaiter’s Kantorovich theorem on Riemannian manifolds [7, Theorem 3.2] (see also Theorem 3.4 below) immediately yields a useful result:

**Theorem 3.2** Denote the smallest eigenvalue of \((-\nabla F_\mu(w_0))\) by $\lambda_{\min}$ and suppose it is greater than $0$. Let $w_k$ be the iterates of the usual Newton iteration (11) or the ones obtained from the shifted Newton iteration (16).

Suppose that the following Newton-Kantorovich condition is satisfied:
\[
q := 4 |F_\mu(w_0)|_g / \lambda_{\min}^2 < 1.
\]
Then there exists a conformal barycenter $w_\ast(\mu)$ of $\mu$, the measure $\mu_\ast := (\sigma_{w_\ast(h)})_\# \mu$ satisfies $\int_{\omega} \mu_\ast = 0$, and the iterates $w_k$ and $\mu_k$ converge with
\[
d_g(w_k, w_\ast(\mu)) \leq \frac{1}{2} \lambda_{\min} q^{(2k)} \quad \text{and} \quad W_p(\mu_k, \mu_\ast) \leq \lambda_{\min} q^{(2k)},
\]
where $W_p$ denotes the $p$-Wasserstein distance for $p \in [1, \infty]$.

The following is a direct consequence of this theorem:

**Corollary 3.3** Let $H = \mathbb{R}^d$ and denote the Euclidean norm by $|\cdot|$. Suppose $\mu = \sum_{i=1}^{n} \omega_i \delta(x_i)$ with $\omega_i > 0$, $\sum_{i=1}^{n} \omega_i = 1$. Denote by $\lambda_{\min}$ the smallest eigenvalue of the matrix $A := \text{id}_{\mathbb{R}^d} - \sum_{i=1}^{n} (x_i x_i^T) \omega_i$ and by $w_{\text{cm}} = \sum_{i=1}^{n} \omega_i x_i$ the center of mass of $\mu$. Suppose that $q := 4 |w_{\text{cm}}| / \lambda_{\min}^2 < 1$ and $w_0 := 0$. Then for any $\varepsilon > 0$, each of the algorithms (11) and (16) reduces the hyperbolic distance between $w_k$ and the conformal barycenter $w_\ast(\mu)$ and the $p$-Wasserstein distance between $\mu_k$ and the centralized measure $\mu_\ast$ to less than $\varepsilon$ in at most $k \leq \lceil \log_2 |\log_2 \varepsilon| - \log_2 |\log_2 q| \rceil$ iterations. For this, they require $O(k(d^2 n + d^3))$ time and $O(d n + d^2)$ memory.

**Proof.** The first part of the corollary comes from plugging definitions into Theorem 3.2 for $w_0 = 0$: We observe that $F_\mu(0) = \frac{1}{2} w_{\text{cm}}$, $|F_\mu(0)|_g = 2 |F_\mu(0)| = |w_{\text{cm}}|$, and $-\nabla F_\mu(0) = \text{id}_{\mathbb{R}^d} - \sum_{i=1}^{n} (x_i x_i^T) \omega_i$. Hence Theorem 3.2 implies the error bounds (18). By taking logarithms, we see that $\lambda_{\min} q^{(2k)} < \varepsilon$ as soon as $k \geq \log_2 |\log_2 (\varepsilon)| - \log_2 |\log_2 q|$.

In each iteration, we have (i) to compute either $F_\mu(w_k)$ or $\mu_k$ and $F_\mu(0)$ which both take $O(d n)$ time; (ii) to compute $\nabla F_\mu(w_k)$ or $\nabla F_\mu(0)$ which both take $O(d^2 n)$ time; (iii) solve a linear equation with a matrix of size $d \times d$ which can be performed in $O(d^3)$ time. We require $O(d n)$ memory for storing $\mu$ and $\mu_k$ and $O(d^2)$ memory to store the $d \times d$ matrix and for computing its inverse. \(\square\)
We now turn to the proof of Theorem 3.2. We first state a (slightly modified) version of Last, the point \( w \). Let \( \gamma \) be a geodesic in \( B \) and let \( \omega \) be a constant \( a \). T \( T \) is invertible and its inverse is a bounded operator on the Hilbert space \( H \). required completeness conditions explicitly, but in localized form. We will use \( |·|_g \) for the norm given by \( g \) for vectors on \( T_s M \) and \( |||·|||_g \) for the corresponding operator norm \(^1\) for linear maps \( A: T_s M \rightarrow T_s M \).

**Theorem 3.4** Let \( (M, g) \) be a (not necessarily finite-dimensional) Riemannian manifold, let \( U \subset M \) be an open set such that \( U \) is geodesically convex\(^2\) and \( \bar{U} \) is complete with respect to the geodesic distance \( d_u \). Let \( F \in C^1(\bar{U}; T M) \) be a continuously differentiable vector field and let \( w_0 \in U \) be a given point where the covariant derivative \( \nabla F(w_0) : T_{w_0} M \rightarrow T_{w_0} M \) is invertible and its inverse is a bounded operator on the Hilbert space \( T_{w_0} M \). Suppose that there are constants \( a > 0, b \geq 0 \), and \( L \geq 0 \) such that

\[

\|\nabla F(w_0)\|_g \leq a \quad \text{and} \quad \|\nabla F(w_0)^{-1} F(w_0)\|_g \leq b

\]

(19)

and \( \nabla F \) obeys the Lipschitz condition that for each two \( z, w \in U \), one has

\[

\|P_{\gamma}(\nabla F(w)) - \nabla F(z)\|_g \leq L d_g(w, z),

\]

(20)

where \( P_{\gamma} : \text{End}(T_w \bar{B}) \rightarrow \text{End}(T_z \bar{B}) \) is the parallel transport along the (unique) minimizing geodesic \( \gamma \) from \( w \) to \( z \) in \( U \). Further, suppose we define the auxiliary constants

\[

r := \frac{1}{a L}, \quad q := 2 a b L < 1, \quad \text{and} \quad r_- := r (1 - \sqrt{1 - q}),

\]

and that we have \( B(w_0; r_-) \subset U \).

Then the sequence of Newton iterates \( w_{k+1} := \exp_{w_k} ( - \nabla F(w_k)^{-1} F(w_k) ) \) is well-defined and contained in \( B(w_0; r_-) \) for all \( k \). They converge to some \( w_s \in \bar{B}(w_0; r_-) \) with \( F(w_s) = 0 \) and

\[

d_g(w_k, w_s) \leq \frac{2 b}{q} q^{(2k)} = \frac{1}{a L} q^{(2k)}.

\]

Last, the point \( w_s \) is the unique zero of \( F \) in \( \bar{B}(w_0; r_-) \).

---

\(^1\)For a linear operator \( A: T_s M \rightarrow T_s M \), it is given by \( ||A||_g = \sup_{w \in T_s M, ||u||_g} |A u|_g / ||u||_g \).

\(^2\)This means that every two points \( x, y \in U \) are connected by a unique length-minimizing geodesic that is contained in \( U \). In particular that guarantees that for every open ball \( B(x, \rho) \subset U \), the Riemannian exponential map \( \exp_x : \{ u \in T_x M | ||u||_g < \rho \} \rightarrow U \) is a well defined diffeomorphism onto \( B(x, \rho) \).
To apply Theorem 3.4 to $F_\mu$, we start by proving that $\nabla F_\mu$ obeys indeed a Lipschitz condition with $L = 2$.

**Proposition 3.5** For any $w, z \in \mathbb{B}$, let $\gamma : [0,1] \to \mathbb{B}$ be the unique minimizing geodesic joining $z$ and $w$ with $\gamma(0) = z$ and $\gamma(1) = w$ and let $P_{\gamma,a,b} : \text{End}(T_{\gamma(a)} \mathbb{B}) \to \text{End}(T_{\gamma(b)} \mathbb{B})$ be the parallel transport along $\gamma$ from $\gamma(a)$ to $\gamma(b)$. Then

$$\|P_{\gamma,1,0}(\nabla F_\mu(w)) - \nabla F_\mu(z)\|_g \leq 2 \, d_\mu(z, w).$$

**Proof.** By the fundamental theorem of calculus and the properties of parallel transport, we have for a general tensor field $A \in C^1(\mathbb{B}; \text{End}(T \mathbb{B}))$ that

$$P_{\gamma,1,0}(A(w)) - A(z) = \int_0^1 \frac{d}{dt}(P_{\gamma,t,0}(A(\gamma(t)))) \, dt = \int_0^1 P_{\gamma,t,0}(\nabla(\gamma)(A)) \, dt.$$

Since $P_{\gamma,t,0}$ is an isometry, we obtain

$$\|P_{\gamma,1,0}(A(w)) - A(z)\|_g \leq \int_0^1 \|P_{\gamma,t,0}(\nabla(\gamma)(A))\|_g \, dt = \int_0^1 \|\nabla(\gamma)(A)\|_g \, dt \leq \int_0^1 \|\nabla(\gamma(\gamma(t)))\|_g \, dt \leq \|\nabla A\|_{L^\infty} \, d_\mu(z, w),$$

where $\|\cdot\|_{L^\infty}$ denotes the supremum-norm with respect to $g$. Thus, for $A = \nabla F_\mu$, it suffices to show that its covariant derivative $\nabla \nabla F_\mu$ is uniformly bounded by 2. Since $\|\nabla \nabla F_\mu(w)\|_g \leq \int_0^1 \|\nabla \nabla V_x(w)\|_g \, dt \leq \|\nabla \nabla V_x\|_{L^\infty}$, it suffices to show that $\nabla \nabla V_x$ is uniformly bounded by 2 which is what we do next. Abbreviate $\alpha_x := V_x^b$. Because $b$ is covariantly constant and isometric, we have $\|\nabla \nabla \alpha_x\|_{L^\infty} = \|\nabla \nabla V_x\|_{L^\infty}$. We deduce from (14) that $\nabla \alpha_x = \alpha_x \otimes \alpha_x - g$. The Leibniz rule yields

$$\nabla X \nabla \alpha_x = \nabla X (\alpha_x \otimes \alpha_x - g) = (\nabla X \alpha_x) \otimes \alpha_x + \alpha_x \otimes (\nabla X \alpha_x) = (\alpha_x(X) \alpha_x - g(X, \cdot)) \otimes \alpha_x + \alpha_x \otimes (\alpha_x(X) \alpha_x - g(X, \cdot)).$$

So with respect to the operator norm on $T^* \mathbb{B} \otimes T^* \mathbb{B} \otimes T^* \mathbb{B}$ given by

$$\|B\|_g := \sup_{X,Y,Z\in T^*_w \mathbb{B} \setminus \{0\}} \frac{|B(X,Y,Z)|}{\|X\|_g \|Y\|_g \|Z\|_g} \quad \text{for} \quad B \in T^*_w \mathbb{B} \otimes T^*_w \mathbb{B} \otimes T^*_w \mathbb{B},$$

we have the following uniform bound for the second derivative of the director field:

$$\|\nabla \nabla V_x\|_{L^\infty} = \|\nabla \nabla \alpha_x\|_{L^\infty} \leq 2 \|\alpha_x\|_g - \|\alpha_x \otimes \alpha_x\|_g \leq 2. \quad \square$$

We are now ready to prove Theorem 3.2.

**Proof.** To apply Theorem 3.4, we must find $a$, $b$ and $L$ so that (19) and (20) are satisfied. We have $\|\nabla F(w_0^{-1})\|_g = \|\nabla F(w_0)^{-1}\|_g = 1/\lambda_{\text{min}}$, where $\lambda_{\text{min}}$ is the minimum eigenvalue of $-\nabla F(w_0)$. Thus we can set $a = 1/\lambda_{\text{min}}$. We can then choose $b$ and check that (19) is satisfied by letting

$$b = |F(w_0)|_g/\lambda_{\text{min}} = a \|F(w_0)\|_g = \|\nabla F(w_0)^{-1}\|_g \|F(w_0)\|_g \geq |\nabla F(w_0)^{-1} F(w_0)|_g.$$

Since $(\mathbb{B}, g)$ is geodesically convex and complete, we may choose $U = \mathbb{B}$. Proposition 3.5 shows that (20) is satisfied with $L = 2$. The rest follows from plugging these constants into the statement of Theorem 3.4. \qed
Figure 4: Empirical distribution functions of $q = 4|w_{cm}|/\lambda_{\min}^2$ for various values of $n$, derived from samples of size $N = 10^6$. Here, we fixed the weights $\omega_1 = \cdots = \omega_n = \frac{1}{n}$ and sampled the point cloud $x$ uniformly from $(\mathbb{S}^2)^n$. As can be seen, the chance of $q$ being close to or greater than 1 rapidly decays with increasing $n$. For example, for $n = 64$, the empirical probability of $P(q > 0.99)$ was lower than 0.03 percent.

4 How often do the Newton-Kantorovich conditions hold?

Having established in Corollary 3.3 that if the Newton-Kantorovich conditions hold, then the conformal barycenter is easy to find, we now turn to the question of how often we find ourselves in these favorable circumstances. Experimentally, the answer is “almost always”, even for surprisingly small values of $n$ (see Figure 4). Theoretically, the answer is “on all but an exponentially small fraction of the space of possible measures”. We now prove this result, though we won’t try to be very sharp in our estimates.

**Theorem 4.1** Suppose $H = \mathbb{R}^d$, $\mu = \sum_{i=1}^n \omega_i \delta(x_i)$ with $\omega_i > 0$, $\sum_{i=1}^n \omega_i = 1$. Let $\lambda_{\min}$ be the smallest eigenvalue of $\text{id}_{\mathbb{R}^d} - \sum_{i=1}^n (x_i x_i^T) \omega_i$, let $w_{cm} = \sum_{i=1}^n \omega_i x_i$ be the center of mass of $\mu$, and let $w_*$ be the conformal barycenter of $\mu$. Let $|.|$ be the standard norm on $\mathbb{R}^d$. Suppose we let $\rho_i = n \omega_i$ be the relative weight of each $\delta(x_i)$, and have $\rho_{\max}$ be the maximum of these weights.

Consider the space $(\mathbb{S}^{d-1})^n$ of possible measures $\mu$ of this type with given weights $\omega_i$. For each $d \geq 2$ there is a universal constant $C(d) > 1$ so that the Newton-Kantorovich condition $q = 4|w_{cm}|/\lambda_{\min}^2 \leq 1/2$ holds on at least the fraction

$$P(q \leq \frac{1}{2}) \geq 1 - 2d C(d)^{-\rho_{\max}}$$

of the space (by volume). We may take the constant $C(d) = \exp(\frac{3}{2}(1 - \frac{1}{d})^4 - \frac{1}{12+\sqrt{d}})$, while $C(2) \geq \exp(\frac{1}{400})$ and $C(3) \geq \exp(\frac{1}{125})$, respectively.

**Proof.** We know that

$$P(q \leq \frac{1}{2}) = P(2|w_{cm}| < \lambda_{\min}^2) \geq P(2|w_{cm}| \leq \frac{4}{3}(1 - \frac{1}{d})^2 \text{ and } \lambda_{\min}^2 \geq \frac{4}{3}(1 - \frac{1}{d})^2 \leq \lambda_{\min}^2)$$

Applying the union bound, we have

$$P(q \leq \frac{1}{2}) \geq 1 - P(|w_{cm}| > \frac{4}{10}(1 - \frac{1}{d})^2) - P(\lambda_{\min} < \sqrt{\frac{4}{3}(1 - \frac{1}{d})})$$

(22)
Therefore, we must separately bound the probability that $|w_{cm}|$ is large and that $\lambda_{\min}$ is small. Proposition 14 in [8] uses Bernstein’s inequality to prove (for $t > 0$)

$$P(|w_{cm}| > t) \leq d \exp\left(-\frac{3n^2}{2t\rho_{\max}\sqrt{d} + 6(1 + n^2 \text{Var} \omega_i)}\right)$$

Using the Bhatia-Davis inequality, we can overestimate $n^2 \text{Var} \omega_i \leq \rho_{\max} - 1$ to simplify the bound to

$$P(|w_{cm}| > t) \leq d \left(\exp\left(\frac{3t^2}{2t\sqrt{d} + 6}\right)\right)^{-\frac{n}{\rho_{\max}}}.$$

Substituting in the value for $t$ given in (22), the function inside the exponential becomes a complicated algebraic function of $d$ which can be underestimated (for $d \geq 2$) by $C(d)$.

For the second part, we will use Proposition 15 in [8] (noting that the sign of the inequality is wrong in the statement of the Proposition; the proof is correct), which uses the matrix Bernstein inequality to prove (for $t > 0$) that

$$P(\lambda_{\min} < (1 - \frac{1}{d}) - t) \leq d \exp\left(-\frac{d}{d - 1} \cdot \frac{3d^2n^2}{2td\rho_{\max} + 6(1 + n^2 \text{Var} \omega_i)}\right)$$

Using the Bhatia-Davis inequality as before, we simplify the bound to

$$P(\lambda_{\min} < (1 - \frac{1}{d}) - t) \leq d \left(\exp\left(\frac{3d^2t^2}{2(d - 1)(3 + dt)}\right)\right)^{-\frac{n}{\rho_{\max}}}.$$

Setting $t = (1 - \frac{1}{d})(1 - \sqrt{4/5})$ to make the left hand side match the corresponding term in (22), we again get a complicated algebraic expression in $d$ which can again be underestimated by $C(d)$. The estimates of $C(2)$ and $C(3)$ are numerical. $\square$

We see that while the largest (relative) mass $\rho_{\max}$ and the dimension $d$ both affect our bound on the fraction of the space of measures where the Newton-Kantorovich conditions hold, this fraction always converges to 1 exponentially quickly in $n$.

5 When the Newton-Kantorovich conditions don’t hold

In extreme cases, e.g., when $\mu_k$ is very concentrated around the two endpoints of a single geodesic, $\nabla F_{\mu_k}(0)$ may be very ill-conditioned (see Lemma 5.6) so that the Newton-Kantorovich conditions may not be satisfied. This leads to an oversized search direction $u_{k+1}$ and the need for backtracking. We can mend this problem by utilizing a regularized variant of Newton’s method (see [15]). Instead of $u_k$ from (16), one may employ the following search direction:

$$u_k := -(\nabla F_{\mu_k}(0) - \alpha |F_{\mu_k}(0)|_g^2 \text{id}_H)^{-1}F_{\mu_k}(0) \quad \text{with} \quad \alpha \geq 0. \quad (23)$$

However, to guarantee convergence, we will have to explicitly account for the possibility of backtracking. We do this in Theorem 5.9, which gives an algorithm which converges under the most general input conditions possible. To prepare for the proof of the theorem, we will now recast the search for the conformal barycenter as a convex optimization problem and study the behavior of the objective function.
5.1 Potentials

We will now find (hyperbolic) potential functions $\psi$ for $V^s$ and $\Psi^s$ for $F^s$; that is, functions $\psi^s: B \to \mathbb{R}$ so that $V^s(w) = -\text{grad}_g(\psi^s)(w)$ and $F^s = -\text{grad}_g(\Psi^s)(w)$. The unique potential for $V^s$ that is gauged to $\psi^s(0) = 0$ can be computed as follows:

$$\psi^s(w) = \int_0^1 \langle -V^s(tw), \frac{d}{dt}(tw) \rangle_g \, dt = \log \left( \frac{|w|^2}{1-|w|^2} \right).$$ (24)

The potential $\Psi^s$ for $F^s$ with the gauge $\Psi^s(0) = 0$, is given by

$$\Psi^s(w) := \int_{\mathbb{S}} \psi^s(w) \, d\mu(x) = \int_{\mathbb{S}} \log \left( \frac{|w|^2}{1-|w|^2} \right) \, d\mu(x).$$ (25)

As pointed out by Douady and Earle in [1, Section 11], $\exp(-\psi^s(w))$ is proportional to the Poisson kernel for $d = 2$. However, for general $d > 2$, $\exp(-\psi^s(w))$ is not harmonic in $w$. With (14), we have

$$\text{Hess}_g(\Psi^s) = \int_{\mathbb{S}} \text{Hess}_g(\psi^s) \, d\mu(x) = -\int_{\mathbb{S}} \nabla V^s \, d\mu(x) = g - \int_{\mathbb{S}} V^s_v \otimes V^s_v \, d\mu(x).$$ (26)

Because $V^s$ is a unit vector field, $\text{Hess}_g(\Psi^s)(w)$ is positive semi-definite and the spectrum of $\text{Hess}_g(\Psi^s)(w)$ is contained in $[0, 1]$. In particular, $\Psi^s$ is convex. Thus the conformal barycenters of $\mu$ coincide with the minimizers of the potential $\Psi^s$. Moreover, we observe that $\text{Hess}_g(\Psi^s)(w)$ is not definite if and only if $\text{supp}(\mu) \subset \{x_+, x_-\}$ and if there is a geodesic that contains $w$, $x_+$, and $w_-$. This motivates the following terminology that we borrow from Kapovich and Millson [3].

**Definition 5.1** A measure $\mu$ on $\mathbb{S}$ is **stable** if no point $x \in \mathbb{S}$ supports half (or more) of the mass of $\mu$, **nice semi-stable** if it is stable or one point $x_+$ supports half the mass of $\mu$ and another point $x_-$ supports half of the mass of $\mu$, **semi-stable** if it is nice semi-stable or if one point $x_+$ supports exactly half the mass of $\mu$, and **unstable** otherwise.

In particular, $\Psi^s$ is strictly convex if $\mu$ is stable. Since the critical points of $\Psi^s$ coincide with the conformal barycenters of $\mu$, this implies that each stable $\mu$ has at most one conformal barycenter. Indeed, it is known that the conformal barycenter of a stable $\mu$ exists (see [1] and Lemma 2.9 in [3]). Also each nice semi-stable $\mu$ has conformal barycenters, but in this case they are not unique as every point on the geodesic joining $x_+$ and $x_-$ is a conformal barycenter. And all other $\mu$ have no conformal barycenter.

For the numerical optimization of $\Psi^s$ and in particular for Newton’s method, it would be most desirable if $\Psi^s$ where **uniformly** convex in the sense that there is a fixed constant $c > 0$ such that $\text{Hess}_g(\Psi^s)(w) \geq c \, g|_{w}$ holds for all $w \in B$. But as we will see soon in Proposition 5.3, such a constant cannot exist, at least globally. Instead, we show that such a constants exists for every ball $B(w_0; r) \subset B$ of finite radius and we give a lower bound for this constant (see Lemma 5.6). This lower bound is in terms of a quantitative measurement of stability for $\mu$ (with respect to the point $w_0$) that we have to develop next. To this end, we introduce the notion of a “viewing cone”.

15
Figure 5: (a) shows a symmetric viewing cone $A(w_1, X; \delta) \cup A(w_1, -X; \delta)$, with the center points $\lim_{t \to \infty} \exp(t X) = x_+$ and $\lim_{t \to -\infty} \exp(t X) = x_-$. In (b), we see that the same subset of $\mathbb{S}$ is also the union of two viewing cones $A(w_2, Y_+; \beta_+) \cup A(w_2, Y_-; \beta_-)$ when viewed from $w_2$. However, as can be seen in (c), the direction vectors $Y_\pm$ are not antipodal to each other, the center points $y_\pm$ are not the same as $x_\pm$ and the angles $\beta_\pm$ are neither equal to each other nor to $\delta$.

5.2 Viewing cones and the Hessian of the potential

Definition 5.2 For $w \in \mathbb{B}$, $X \in \text{UT}_w \mathbb{B}$, and $\delta \in [0, \pi]$, we define the viewing cone (see Figure 5, (a)).

$$A(w, X; \delta) := \{ y \in \mathbb{S} \mid \langle X, V_y(w) \rangle_g \geq \cos(\delta) \}.$$  

From the transformation rules (4) for $V_x$, it follows immediately that the viewing cones transform as follows under the Möbius transformation $\varphi \in \text{Aut}(\mathbb{B}, g)$:

$$\varphi (A(w, X; \delta)) = A(\varphi(w), d\varphi(w) X; \delta).$$  

(27)

This notion of viewing cone allows us to provide bounds for the first eigenvalue of the Hessian of $\Psi_\mu$ in terms of concentration of $\mu$ in a symmetric pair of viewing cones of given angle:

Proposition 5.3 Let $\beta > 0$ and define

$$a(w) := \inf_{X \in \text{UT}_w \mathbb{B}} \mu(\mathbb{S} \setminus (A(w, X; \beta) \cup A(w, -X; \beta))).$$

Then the smallest eigenvalue $\lambda_\mu(w)$ of $\text{Hess}_g(\Psi_\mu)(w)$ with respect to $g|_w$ satisfies

$$a(w) \sin^2(\beta) \leq \lambda_\mu(w) \leq \sin^2(\beta) + a(w) \cos^2(\beta).$$

So if $a(w) > 0$, the condition number of the Hessian is bounded by $\kappa(\text{Hess}_g(\Psi_\mu)) \leq \frac{1}{a(w)} \csc^2(\beta)$.

In particular this shows that $\lambda_\mu(w)$ can be arbitrarily small when almost all the measure of $\mu$ is supported in a single double viewing cone $A(w, X; \beta) \cup A(w, -X; \beta)$ with small opening angle $\beta$. Indeed, if $\mu$ is supported in $A(w, X; 0) \cup A(w, -X; 0)$, then $\lambda_\mu(w) = 0$. 

16
Proof. Fix $\beta > 0$ and $w \in \mathcal{B}$. For $X \in \mathbb{UT}_w \mathcal{B}$, define the symmetric double cone $A(X) := A(w, X; \beta) \cup A(w, -X; \beta)$ and $b(X) := \mu(\mathcal{S} \setminus A(X))$. Using (26), we have

$$\text{Hess}_g(\Psi \mu)(w)(X, X) = 1 - \int_{\mathcal{S} \setminus A(X)} \langle X, V_y(w) \rangle_g \, d\mu(y) - \int_{A(X)} \langle X, V_y(w) \rangle_g \, d\mu(y).$$

For $y$ outside of $A(X)$, we have the bound $\langle X, V_y(w) \rangle_g^2 \leq \cos^2(\beta)$, while we only know that $\langle X, V_y(w) \rangle_g^2 \leq 1$ in case of $y \in A(X)$. Thus we obtain

$$\text{Hess}_g(\Psi \mu)(w)(X, X) \geq 1 - \int_{\mathcal{S} \setminus A(X)} \cos(\beta)^2 \, d\mu(y) - \int_{A(X)} \, d\mu(y)$$

$$= 1 - b(X) \cos(\beta)^2 - (1 - b(X)) = b(X) \sin^2(\beta) \geq a(w) \sin^2(\beta).$$

It is well-known that $\lambda_\mu(w) = \inf_{X \in \mathbb{UT}_w \mathcal{B}} \text{Hess}_g(\Psi \mu)(w)(X, X)$, so $\lambda_\mu(w)$ is the greatest lower bound for $\{\text{Hess}_g(\Psi \mu)(w)(X, X) \mid X \in \mathbb{UT}_w \mathcal{B}\}$. We have just proved that $a(w) \sin^2(\beta)$ is some lower bound for the same set. Thus we obtain $\lambda_\mu(w) \geq a(w) \sin^2(\beta)$. Since the maximum eigenvalue of $\text{Hess}_g(\Psi \mu)$ is at most 1, our estimate of the condition number $\kappa(\text{Hess}_g(\Psi \mu))$ follows immediately.

For $y$ inside of $A(X)$, we know $\langle X, V_y(w) \rangle_g^2 \geq \cos^2(\beta)$, while outside of $A(x)$ we know only $\langle X, V_y(w) \rangle_g^2 \geq 0$. Thus

$$\lambda_\mu(w) \leq \text{Hess}_g(\Psi \mu)(w)(X, X) \leq 1 - 0 - \int_{A(X)} \cos^2(\beta) \, d\mu(y)$$

$$\leq 1 - (1 - b(X)) \cos^2(\beta) = \sin^2(\beta) + b(X) \cos^2(\beta).$$

As before we now know that $\lambda_\mu(w)$ is a lower bound for $\{|\sin^2(\beta) + b(X) \cos^2(\beta)| \mid X \in \mathbb{UT}_w \mathcal{B}\}$. Since $\sin^2(\beta) + a(w) \cos^2(\beta)$ is the greatest lower bound for the same set, we have $\lambda_\mu(w) \leq \sin^2(\beta) + a(w) \cos^2(\beta)$, as desired. $\square$

5.3 Quantitative stability

We now want to show that Proposition 5.3 can be applied to any stable Borel probability measure $\mu$ on $\mathcal{S}$. To that end, we will first show

**Lemma 5.4** Let $\mu$ be a Borel probability measure on the complete metric space $(M, d)$, such that no singleton has measure 1/2 or more, i.e. $\mu(\{x\}) < 1/2$ for all $x \in M$ or $\mu$ is stable. Then there are $r > 0$ and $\epsilon > 0$ such that for all $x \in M$ one has $\mu(\bar{B}(x; r)) \leq (1 - \epsilon)/2$, where $\bar{B}(x; r)$ is the closed ball of radius $r$ around $x$.

Recall that the viewing cone $A(0, x/2; \delta) \subset \mathcal{S}$ is just a closed spherical ball around $x \in \mathcal{S} \cong \mathbb{UT}_0 \mathcal{B}$ of radius $\delta$. Thus, Lemma 5.4 shows that every stable Borel probability measure on $\mathcal{S}$ has to satisfy the following property for $w = 0$:

$$\text{There are } \epsilon > 0, \delta > 0 \text{ s.t. } \mu(A(w, X; \delta)) \leq \frac{1 - \epsilon}{2} \text{ holds for all } X \in \mathbb{UT}_w \mathcal{B}. \quad (28)$$

Now Möbius transformations map stable measures to stable measures, which immediately yields:
Lemma 5.5 Let \( \mu \) be a Borel probability measure on \( \mathbb{S} \). Then the following three statements are equivalent:

1. \( \mu \) is stable;
2. condition (28) is satisfied for just one point \( w \in \mathbb{B} \); and
3. a condition like (28) holds for each point \( w \in \mathbb{B} \) (with varying \( \varepsilon \) and \( \delta \)).

We now turn to the proof of Lemma 5.4:

Proof. Assume that this were false. Then for each \( n \in \mathbb{N} \) there is an \( x_n \in M \) with \( \mu(B(x_n; 4^{-n})) > (1 - 4^{-n})/2 \). We abbreviate \( \Omega_n := \overline{B}(x_n; 4^{-n}) \) and note that each \( \Omega_n \) has measure greater than 1/3.

Claim: There is a subsequence \( (\eta_k)_{k \in \mathbb{N}} \) such that \( \Omega_{n_k} \cap \Omega_{m_k} \neq \emptyset \) for each \( k \) and each \( m \geq k \).

We thin out the sequence \( a_1 := (\Omega_n)_{n \in \mathbb{N}} \) recursively so obtain sequences \( a_1, a_2, a_3 \ldots \) such that \( a_{k+1} \) is a subsequence of \( a_k \) and such that \( a_{k,1} \) of each \( a_k \) has nontrivial intersection with all \( a_{k,i}, i \geq 1 \). Then \( (a_{k,i})_{i \in \mathbb{N}} \) is the subsequence of \( (\Omega_n)_{n \in \mathbb{N}} \) that we are looking for. To this end we consider \( k \in \mathbb{N} \) and the subsequence \( a_k \) of \( (\Omega_n)_{n \in \mathbb{N}} \). It suffices to show that this sequence must contain an \( a_{k,i_0} \) that is intersected by infinitely many of the \( a_{k,j}, j \geq i_0 \) because we can then define \( a_{k+1} \) as the subsequence of \( a_k \) that contains \( a_{k,i_0} \) and all the \( a_{k,j}, j \geq i_0 \) that intersect \( a_{k,i_0} \). Indeed, such an \( a_{k,i_0} \) does exist: Either \( a_{k,1} \) intersects infinitely many \( a_{k,j} \) or not. If not, then there is a first \( a_{k,i} \) that is disjoint from \( a_{k,1} \). For every \( j \geq i \) we have \( \mu(a_{k,1} \cup a_{k,j}) + \mu(a_{k,i}) = \mu(a_{k,1}) + \mu(a_{k,i}) + \mu(a_{k,j}) > 1 \). This means that each \( a_{k,j}, j \geq i \) must intersect \( a_{k,1} \) or \( a_{k,i} \), since the total measure of \( \mu \) is 1. Thus at least one of \( a_{k,1} \) and \( a_{k,i} \) is intersected by infinitely many \( a_{k,j} \). Altogether, this proves the claim.

Having found the sequence \( (a_k)_{k \in \mathbb{N}} \), we observe that \( (x_n)_{n \in \mathbb{N}} \) must be a Cauchy sequence. Indeed, for each \( N \in \mathbb{N} \) and all \( j \geq i \geq N \), the balls \( \overline{B}(x_n; 4^{-n}) \) and \( \overline{B}(x_m; 4^{-m}) \) have nontrivial intersection, thus we have \( d(x_n, x_m) \leq 2 \cdot 4^{-n} \). Thus there exists a limit point \( x \) of \( (x_n)_{n \in \mathbb{N}} \) and we have \( d(x, x_n) \leq 2 \cdot 4^{-n} \). In particular, this implies \( B(x; 4 \cdot 4^{-n}) \supset \overline{B}(x_n; 4^{-n}) \) and, because every finite Borel measure on metric spaces is outer regular, we have

\[
\mu(\{x\}) = \lim_{k \to \infty} \mu(\overline{B}(x; 4 \cdot 4^{-n})) \geq \lim_{k \to \infty} \mu(\overline{B}(x_n; 4^{-n})) \geq \lim_{k \to \infty} \frac{1 - 4^{n_k}}{2} = \frac{1}{2}
\]

which is a contradiction to the stability of \( \mu \). So the initial assumption must be wrong and we have proven the lemma.

\[\square\]

5.4 Uniform convexity of the potential

We now prove that the potential \( \Psi_\mu \) of a stable Borel probability measure \( \mu \) is uniformly convex on balls of finite radius and provide an explicit bound on the lowest eigenvalue of the Hessian.

Lemma 5.6 Let \( w_0 \in \mathbb{S} \) and suppose that \( \mu \) satisfies (28) for \( w = w_0 \). Let \( \lambda_\mu(w) \) denote the smallest eigenvalue of \( \text{Hess}(\Psi_\mu)(w) \). Then for all \( w \in \mathbb{B} \) we have

\[\lambda_\mu(w) \geq \varepsilon \sin^2(\delta) \exp(-2d_\delta(w_0, w)).\]

In particular, \( \Psi_\mu \) is uniformly convex and the condition number of \( \text{Hess}_\delta(\Psi_\mu)(w) \) is uniformly bounded on hyperbolic balls around \( w_0 \) of finite radius.

18
Proving Lemma 5.6 will require us to investigate a further property of viewing cones: Each viewing cone \( A(w, X, \delta) \) is a closed ball with respect to the angular metric on the sphere \( \mathbb{S} \). This can be most easily seen by applying the Möbius transformation \( \varphi \in \text{Aut}(\mathbb{B}, g) \) and by recalling that the induced Möbius transformation \( \varphi : \mathbb{S} \to \mathbb{S} \) maps spherical balls to spherical balls. Thus every viewing cone \( A(w_1, X_1, \delta_1) \) at point \( w_1 \) coincides with another viewing cone \( A(w_2, X_2, \delta_2) \) at point \( w_2 \). However, the center and angle of the cone change as we move from \( w_1 \) to \( w_2 \) (see Figure 5). We now estimate this change of the angles:

**Lemma 5.7** Let \( w_1 \in \mathbb{B}, X_1 \in \text{UT}_w \mathbb{B}, \) and \( \delta_1 \in (0, \pi) \). Then for each \( w_2 \in \mathbb{B} \) there exist unique \( X_2 \in \text{UT}_w \mathbb{B} \) and \( \delta_2 \in (0, \pi) \) such that the viewing cones \( A(w_1, X_1; \delta_1) \) and \( A(w_2, X_2; \delta_2) \) coincide. Further, we have

\[
\exp(-d_g(w_1, w_2)) \leq \frac{\sin(\delta_2)}{\sin(\delta_1)} \leq \exp(d_g(w_1, w_2)).
\]

**Proof.** By applying the shift transformation \( \sigma_{w_1} \), we may assume that \( w_1 = 0 \). We fix the shift transformation \( \varphi := \sigma_{w_2} : \mathbb{B} \to \mathbb{B} \) which extends to a Möbius transformation of \( \mathbb{S} \). Observe that \( A(0, X_1; \delta_1) \) is a ball on the sphere with center \( x_1 := \lim_{t \to \infty} \exp_0(t X_1) \). Since \( \varphi \) is a Möbius transformation on \( \mathbb{S} \) and because Möbius transformations map balls to balls, also \( \varphi(A(0, X_1; \delta_1)) \) has to be a ball. Thus it can be written as a viewing cone from the origin: \( \varphi(A(0, X_1; \delta_1)) = A(0, Y; \delta_2) \) with some \( Y \in \text{UT}_0 \mathbb{B} \) and \( \delta_2 > 0 \). Letting \( X_2 := \varphi^{-1}(Y) \), (27) shows that

\[
A(w_1, X_1; \delta_1) = A(0, X_1; \delta_1) = \varphi^{-1}(A(0, Y; \delta_2)) = A(\varphi^{-1}(0), d\varphi(y)^{-1}Y; \delta_2) = A(w_2, X_2; \delta_2),
\]

which completes the proof of the first statement of the lemma.

We still need to estimate \( \delta_2 \). Observe that \( w_1 = 0, w_2 \) and \( x_1 \) lie in a common two-dimensional plane \( E \subset H \). Recalling (6), we observe that the shift \( \varphi \) taking \( w_2 \) to \( 0 \) takes any point \( z \in E \) to a linear combination of \( z \) and \( w_2 \); that is, to a point in \( E \). The inverse map \( \varphi^{-1} = \sigma_{-w_2} \) has the same property, so \( \varphi \) maps \( E \cap \mathbb{B} \) (and \( E \cap \mathbb{S} \)) to itself.

An easy computation shows that if \( z_1 \in E \) and \( z_2 \in E^+ \) then \( \varphi(z_1 + z_2) + \varphi(z_1 - z_2) \in E \). Since the ball \( A(0, X_1; \delta_1) \) is centrally symmetric with respect to \( E \), this shows that the ball \( \varphi(A(0, X_1; \delta_1)) = A(0, Y; \delta_2) \) is also centrally symmetric with respect to \( E \), and in particular its center \( y := \lim_{t \to \infty} \exp_0(t Y) = 2Y \) must lie in \( E \).

We can conclude that \( A(0, X_1; \delta_1) \subset \mathbb{S} \) and \( A(0, Y; \delta_2) \subset \mathbb{S} \) are spherical balls centered at points \( x_1 \) and \( y \) in \( E \cap \mathbb{S} \). Thus \( E \) intersects each of the balls in a diameter and \( \varphi \) maps the diameter \( \{x_+, x_-\} \) of \( A(0, X_1; \delta_1) \cap E \) to the diameter \( \{\varphi(x_+), \varphi(x_-)\} \) of \( A(0, Y; \delta_2) \cap E \). Without loss of generality, we may identify \( E \) with \( \mathbb{R}^2 \) and assume that \( X_1 = (1, 0), w_2 = (\cos(\theta), \sin(\theta)) \), and \( x_\pm = (\cos(\pm \delta_1), \sin(\pm \delta_1)) \). Then we have \( \sin^2(\delta_2) = (1 - \cos(2 \delta_2))/2 = (1 - \langle \varphi(x_+), \varphi(x_-) \rangle)/2 \) and a short computation involving (6) and \( \cos(\delta_1 \pm \theta) \geq -1 \) lead to

\[
\sin^2(\delta_2) = \frac{(1 - r)^2 (1 + r)^2 \sin^2(\delta_1)}{(1 - 2r \cos(\delta_1 - \theta) + r^2)(1 - 2r \cos(\delta_1 + \theta) + r^2)} \geq \frac{(1 - r)^2}{(1 + r)^2} \sin^2(\delta_1).
\]
Both $\delta_1$ and $\delta_2$ are contained in the interval $[0, \pi]$ where $\sin$ is nonnegative, thus we may apply the square root; substituting $r = \tanh(d_g(w_1, w_2)/2)$ leads to

$$\sin(\delta_2) \geq \exp(-d_g(w_1, w_2)) \sin(\delta_1).$$

This shows the first inequality stated by the lemma. The second follows from swapping the roles of $w_1$ and $w_2$. \hfill $\square$

We are now ready to prove Lemma 5.6.

**Proof.** Fix $w \in B$ and $X \in UT_w B$. By Lemma 5.7, there are unit tangent vectors $Y_x \in UT_{w_0} B$ and angles $\beta_x > 0$ so that $A(w, \pm X; \beta_x) = A(w_0, Y_x; \delta)$, as shown in Figure 5. Notice that the two vectors $Y_x$ are likely not some pair $\pm Y$ of antipodal vectors. We put $\beta := \min \{\beta_-, \beta_+, \pi/2\}$ and observe that Lemma 5.7 implies $\sin^2(\beta) \geq \exp(-2 d_g(w, w_0)) \sin^2(\delta)$. By the choice of $\beta$, we have $\mu(A(w, \pm X; \beta)) = \mu(A(w_0, Y_x; \delta)) \leq (1 - \epsilon)/2$. Now $0 \leq \beta \leq \pi/2$ implies

$$b(X) := \mu(\mathbb{S} \setminus (A(w, X; \beta) \cup A(w, -X; \beta))) = 1 - \mu(A(w, X; \beta)) - \mu(A(w, -X; \beta)) \geq \epsilon.$$  

Since $X$ was arbitrary, $a(w) = \inf_{X \in UT_w B} b(X) \geq \epsilon$ holds as well. Proposition 5.3 then yields

$$\lambda_\mu(w) \geq a(w) \sin^2(\beta) \geq \epsilon \sin^2(\delta) \exp(-2 d_g(w, w_0)).$$

The largest eigenvalue of $\Hess_g(\Psi_\mu)$ is at most 1, so this provides a uniform estimate on the condition number of this matrix as well. \hfill $\square$

### 5.5 A general bound on the location of the conformal barycenter

Recall that conformal barycenters of $\mu$ coincide with minimizers of the potential $\Psi_\mu$. Thus the following provides an *a priori* bound on the distance that a conformal barycenter (if existent) can have from a given point.

**Lemma 5.8** Let $w_0 \in \mathbb{S}$ and suppose that $\mu$ satisfies (28) for $w = w_0$. Then we have $\Psi_\mu(w) > \Psi_\mu(w_0)$ for each $w$ outside the closed ball $B$ around $w_0$ of radius $r(\epsilon, \delta) := -(2/\epsilon) \log(\sin(\delta)/2)$.

**Proof.** By applying a shift $\sigma_{w_0}$, we may assume without loss of generality that $w_0 = 0$. We recall that $\Psi_\mu$ was chosen so that $\Psi_\mu(0) = 0$. We now prove that $\Psi_\mu$ is strictly positive outside of the closed ball $B := \{w \in B \mid d_g(w, 0) \leq r(\epsilon, \delta)\}$.

For $x \in \mathbb{S}$, put $X := V_x(0) = \frac{1}{2} x$. Denote the unit speed geodesic ray emanating from 0 in direction $X$ by $\gamma_x(r) := \exp_0(r X) = \tanh(r/2) x$ (see (2)). With (24), we obtain for each $y \in \mathbb{S}$ that

$$\psi_{\gamma_x}(r) = \log(\cosh(r) - \cos(\theta) \sinh(r)), \quad \text{where } \cos(\theta) = \langle x, y \rangle.$$  

As a restriction of a convex function to a geodesic, $r \mapsto \psi_{\gamma_x}(r)$ is convex. Further, a direct computation shows that it has slant asymptotes which we may employ as lower bounds:

$$\psi_{\gamma_x}(r) \geq \min \{ r + 2 \log(\sin(\theta/2)), -r + 2 \log(\cos(\theta/2)) \}.$$
For $0 \leq \theta \leq \pi$, $\log(\sin(\frac{\theta}{2}))$ is monotonically increasing, while $\log(\cos(\frac{\theta}{2}))$ is monotonically decreasing. So we have the bounds

$$\psi_{\gamma}(r) \geq \begin{cases} 
    r + 2 \log(\sin(\frac{\delta}{2})), & \text{for } \delta \leq \theta \leq \pi,
    \\
    -r + 2 \log(\cos(\frac{\delta}{2})), & \text{otherwise}.
\end{cases}$$

We abbreviate $a := \mu(A(0, X; \delta)) \leq \frac{1 - \varepsilon}{2}$. Thus, we have for all $r > r(\varepsilon, \delta) > 0$ that

$$\Psi_{\mu}(\gamma_x(\tau_k v_k)) \geq \int_{S \setminus A(0, X; \delta)} \psi_{\gamma}(\gamma_x(r)) d\mu + \int_{A(0, X; \delta)} \psi_{\gamma}(\gamma_x(r)) d\mu + \int_{A(0, X; \delta)} (-r + 2 \log(\cos(\frac{\delta}{2}))) d\mu(y)$$

$$\geq \varepsilon r + 2 \log(\min(\sin(\frac{\delta}{2}), \cos(\frac{\delta}{2}))) \geq \varepsilon r + 2 \log(\sin(\delta)/2) > 0. \quad \Box$$

### 5.6 Main result: A fast, robust, and globally convergent algorithm

We are now prepared to state a damped and regularized version of Newton’s method (11). By performing an appropriate line search, we can guarantee that the method converges for each stable measure and each starting values.

Choosing a regularization parameter $\alpha \geq 0$, we start with some initial guess $w_0 \in \mathbb{B}$ and iteratively define:

$$v_k := -(\nabla F_{\mu}(w_k) - \alpha |F_{\mu}(w_k)|^2 \text{id}_{\tau_k \mathbb{B}})^{-1} F_{\mu}(w_k), \quad (29)$$

$$w_{k+1} = \exp_{w_k}(\tau_k v_k), \quad (30)$$

where the $\tau_k > 0$ are chosen by a line search so that for some given constants $0 < c_1 \leq \frac{1}{2}$ and $c_1 < c_2 < 1$, the following conditions are met along the search line $\gamma_k(t) := \exp_{w_k}(t v_k)$ for the merit function $f_k(t) := \Psi_{\mu}(\gamma_k(t))$:

$$f_k(\tau_k) \leq f_k(0) + c_1 \tau_k f'_k(0) \quad \text{(Armijo condition)} \quad (31)$$

$$f'_k(\tau_k) \geq c_2 f'_k(0). \quad \text{(weak Wolfe condition)} \quad (32)$$

Moreover, we require that

$$\tau_k = 1 \text{ whenever this choice satisfies (31) and (32).} \quad (33)$$

We note that the Armijo and weak Wolfe condition are standard conditions to guarantee global convergence in unconstrained optimization algorithms. It is well-known how to realize these conditions in a line search algorithm (see for example [16, Chapter 3]). Such algorithms only involve evaluation of $f_k(t) - f_k(0)$ and $f'_k(t)$. For an isometry $\varphi \in \text{Aut}(\mathbb{B})$, the potential $\Psi_{\varphi_{\gamma_k}} \circ \varphi$

---

1In our experiments, the choice $\alpha = 1$ turned out to work best.
diffe from $\Psi_\mu$ only by a constant, and so the line search can also be pushed to the origin via the shift transformation $\sigma_{w_k}$. This leads us to the following shifted variant of $(33)$:

$$\mu_k := (\sigma_{w_k})_\# \mu$$

(34)

$$u_k := -(\nabla F_{\mu_k}(0) - \alpha [F_{\mu_k}(0)]^2 g \text{id}_\mu)^{-1} F_{\mu_k}(0)$$

(35)

$$w_{k+1} := \sigma(-w_k, \exp_0(\tau_k u_k)),$$

(36)

where $\tau_k$ is determined as in $(31)$–$(33)$ but with the merit function $f_k$ replaced by $f_k(t) := \Psi_{\mu_k}(\exp_0(t u_k))$.

Next we show that the sequence $(w_k)_{k \in \mathbb{N}}$ created by these algorithms converges to the conformal barycenter for each stable Borel probability measure $\mu$ on $\mathbb{S}$. As a side effect, we obtain a proof of the existence and uniqueness of the conformal barycenter of any stable Borel probability measure. In contrast to the original proof by Douady and Earle [1, Proposition 1], this proof does not rely on the Poincaré-Hopf index theorem and thus also works in nonseparable Hilbert spaces (for which even the Browder-Minty theorem cannot be applied to show existence).

**Theorem 5.9** Suppose $\mu$ is a Borel probability measure on $\mathbb{S}$ which is stable in the sense of Definition 5.1. Then there exists a unique conformal barycenter $w_*(\mu)$ of $\mu$ and the iterates $(w_k)_{k \in \mathbb{N}}$ defined by $(29)$–$(30)$ or $(34)$–$(36)$ converge quadratically to $w_*(\mu)$ in the sense that there is a $C > 0$ (depending on $\mu$) so that

$$\limsup_{k \to \infty} \frac{d_\mu(w_{k+1}, w_*(\mu))}{d_\mu(w_k, w_*(\mu))^2} \leq C.$$

**Proof.** We employ the techniques from [9] utilizing the Riemannian exponential map as retraction. Since $\mu$ is stable, Lemma 5.5 and Proposition 5.3 imply that the operator $\nabla F_{\mu_k}(w) : T_w \mathbb{B} \to T_w \mathbb{B}$ is continuously invertible at each point $w \in \mathbb{B}$. By [9, Proposition 1], there is always a step size $\tau_k$ satisfying $(31)$–$(33)$. Thus the sequence $(w_k)_{k \in \mathbb{N}}$ is well-defined.

The Armijo condition $(31)$ guarantees $\Psi_{\mu_k}(w_k) < \Psi_{\mu_k}(0)$ and thus Lemma 5.8 implies that $(w_k)_{k \in \mathbb{N}}$ stays within a ball $B$ of finite hyperbolic radius. By Lemma 5.6, this in turn implies that there are $0 < \lambda \leq 1 < \infty$ such that the self-adjoint linear operators $A_{\alpha,k} := -\nabla F_{\mu_k}(w_k) + \alpha [F_{\mu_k}(w_k)]^2 g \text{id}_{T_{w_k}\mathbb{B}}$ satisfy the uniform bound

$$\lambda \text{id}_{T_{w_k}\mathbb{B}} \leq A_{\alpha,k} \leq \Lambda \text{id}_{T_{w_k}\mathbb{B}} \quad \text{for all } k \in \mathbb{N}.$$

Notice that $\Psi_{\mu_k}$ is Lipschitz contiuous differentiable. Since it is convex with $\Psi_{\mu}(0) = 0$ and finite slope at $|\text{grad}_g(\Psi_{\mu}(0))|_g < \infty$, it is also bounded from below on the ball $B$. So [9, Corollary 3] implies so-called global convergence in the sense that

$$|F_{\mu_k}(w_k)|_g = |d\Psi_{\mu_k}(w_k)|_g \xrightarrow{k \to \infty} 0.$$

Notice that these eigenvalues are uniformly bounded from below by $\lambda$. So there must be a $k_0$ such that $q_k := 4 |F_{\mu_k}(w_k)|_g^2 \leq 4 |F_{\mu_k}(w_k)|_g^2 \lambda^{-2} < 1$ holds all $k \geq k_0$. For those $k$, the Newton-Kantorovich theorem (for the Newton algorithm started at $w_k$, see Theorem 3.2) implies the
existence of a conformal barycenter. Moreover, since $\Psi_\mu$ is strictly convex, there is exactly one such conformal barycenter $w_*(\mu)$. Hence the estimate from Theorem 3.2 (with $w_0$ replaced by $w_k$) implies

$$d_g(w_k, w_*(\mu)) \leq \frac{1}{2} \lambda_k q_k^{(2^q)} + 2 |F_\mu(w_k)|_g/\lambda_k \leq 2 |F_\mu(w_k)|_g/\lambda \to \infty 0.$$

We are left to show quadratic convergence. Since we know now that the limit point $w_*(\mu)$ exists, this is fairly standard: Because $F_\mu$ is 1-Lipschitz, we obtain that

$$|d\Psi_\mu(w_k) + \text{Hess}(\Psi_\mu)(w_k)(v_k, -)| = |F_\mu(w_k) + A_{0,k} v_k|_g =$$

$$\leq |A_{0,k} v_k + A v_k|_g = \alpha |F_\mu(w_k)|^2 v_k|_g =$$

$$\alpha |A_{0,k} v_k|_2^2 |v_k|_g \leq \alpha (1 + \alpha)^2 |v_k|_g^2 \to 0.$$

Now [9, Proposition 5] shows that the condition (33) enforces $\tau_k = 1$ for all sufficiently large $k$. Finally, Propositions 7 and 8 from [9] imply quadratic convergence of $w_k$ towards $w_*(\mu)$. □

Due to the regularization, the line search is seldom required in practice. In fact, it may cause some problems when run with finite precision: All computations involving the shift transformation suffer a slight loss of precision. So when the slope $f_k(0) = d\Psi_\mu u_k$ is already very close to 0, the Armijo condition may just not be justifiable due to fact that $f_k(\tau_k) = \Psi_\mu(w_{k+1})$ cannot be computed arbitrarily well. Fortunately, this typically happens only when the Newton-Kantorovich condition $q_k < 1$ is already satisfied (here $q_k := 4 |F_\mu(0)|/\lambda^2_k$ and $\lambda_k$ is the smallest eigenvalue of $-\nabla F_\mu(0)$, see also Theorem 3.2). So one is better off by just putting $\tau_k = 1$ and skipping the check for the Armijo condition whenever $q_k < 1$. Putting also $\alpha = 0$ lets the method fall back to Newton’s method and its convergence is then guaranteed by Theorem 3.2. Since one has to compute $\nabla F_\mu(0)$ anyway, computing the smallest eigenvalue does not really increase the complexity of the algorithm when $d$ is small. Moreover, it allows to use the condition

$$q_k < 1 \quad \text{and} \quad 2 |\nabla F_\mu(0)|/\lambda_k < \varepsilon$$

(37)
as reliable stopping criterion and as an a posteriori error bound. Indeed, the residual $|F_\mu(0)|$ is a very bad predictor of the distance between $w_k$ and $w_*(\mu)$: It typically underestimates the distance and it does so by orders of magnitude when the smallest eigenvalue of $\text{Hess}(\Psi_\mu(w_*(\mu)))$ is tiny.

It might also be noteworthy that the Abikoff-Ye iteration is of the form

$$u_k = F_\mu(0) = -\text{grad}(\Psi_\mu(0)) \quad \text{and} \quad w_{k+1} = \sigma(-w_k, \exp_0(\tau_k u_k)) = \sigma(-w_k, 2 u_k)$$

with step size $\tau_k \to 2$, for $k \to \infty$. Thus it is basically the method of steepest descent, and adding a line search as above would also make this method globally convergent (but of course, only with linear convergence rate).
6 Experimental results

We now give some examples of the performance of our methods in practice. We start with the Douady-Earle extension which was the original motivation for studying conformal barycenters (see [1]). Afterwards, we show a couple of examples for polygonal closure which was our initial motivation.

6.1 Douday-Earle extension

**Definition 6.1** Suppose we are given a nonconstant, continuous map $\gamma: \mathbb{S}_1 \to \mathbb{S}_2$ between the unit spheres $\mathbb{S}_1$ and $\mathbb{S}_2$ in the Hilbert spaces $H_1$ and $H_2$. Suppose that $H_1$ is finite-dimensional; then there is a unique rotation-invariant Borel probability measure $\nu_0$ on $\mathbb{S}_1$. For each $z \in \mathbb{B}_1$, one defines the measure $\nu_z := \sigma_z^\# \nu_0$ via pullback along the shift transformation $\sigma_z$. Clearly, this measure has the property that the conformal barycenter of $\nu_z$ is $z$ itself, i.e., $w_*(\nu_z) = z$. Since $\gamma$ is nonconstant and continuous, the measure $\gamma_\# \nu_z$ cannot be concentrated within two single points. So it is stable and there is a unique conformal barycenter $w_*(\gamma_\# \nu_z)$. Thus we may define the **Douady-Earle extension** $E(\gamma): \mathbb{B}_1 \to \mathbb{B}_2$ of $\gamma$ by setting $E(\gamma)(z) := w_*(\gamma_\# \nu_z)$.

By (5), the extension operator $E$ is “conformally natural”. This means that if $\varphi_1$ is a Möbius transformation of (the closure of) $\mathbb{B}_1$ then $E(\gamma \circ \varphi_1) = E(\gamma) \circ \varphi_1$. Further, if $\varphi_2$ is a Möbius transformation of (the closure of) $\mathbb{B}_2$ then $E(\varphi_2 \circ \gamma) = \varphi_2 \circ E(\gamma)$. These properties are illustrated in Figure 6.

One can approximate $E(\gamma)$ by approximating $\nu_0$ by a discrete $n$-point measure $\nu_{0,n}$. If $\gamma$ is a sufficiently smooth and for fixed $z \in \mathbb{B}_1$, the conformal barycenter $E_n(\gamma)(z)$ of $\gamma_\# \sigma_z^\# \nu_{0,n}$ will
converge to $E(\gamma)(z)$ provided that $\nu_{0,n}$ converges in 1-Wasserstein distance to $\nu_0$. Approximating $\nu_0$ is particularly easy if $S_1 = S^1$ is the 1-dimensional sphere: We may choose uniformly distributed quadrature points and put $\nu_{0,n} = \frac{1}{n} \sum_{i=1}^{n} \delta(\cos(\frac{2\pi}{n}i), \sin(\frac{2\pi}{n}i))$. Since $\gamma_# \sigma_n^g \nu_{0,n}$ is also a discrete measure, we may compute its conformal barycenter by the method outlined in Section 5. The reader might find it intriguing to try out the Mathematica routine DouadyEarleExtension provided by the package ConformalBarycenter.m in the electronic supplement. This routine computes $E_n(\gamma)$ of a “piecewise-linear” curve $\gamma$: $S^1 \subset \mathbb{R}^2 \to S^2 \subset \mathbb{R}^3$. It is the very routine that we used to produce Figure 6 and Figure 7.

This application is actually quite challenging: For $z$ close to the boundary of $B_1$, the measure $\nu_z$ on $S_1$ and its pushforward $\gamma_# \nu_z$ on $S_2$ are highly concentrated. This means that the Newton-Kantorovich condition from Theorem 3.2 may be far from being satisfied at the starting point because the Hessian is very degenerate. The method of steepest descent is notorious for having very oscillatory behavior and for not making good progress in such regions. Hence the method of Milnor-Abikoff-Ye should suffer immensely from degenerate Hessians. Indeed, the left hand side of Figure 7 shows that the Milnor-Abikoff-Ye iteration has problems with computing the Douady-Earle extension in the boundary regions of the surfaces. There, the algorithm did not converge even after 1000 iterations. Despite this, the damped, regularized Newton method (with regularization parameter $\alpha = 1$ and initialized with the Euclidean center of mass of $\gamma_# \nu_z$) required typically less than a dozen iterations to decrease the Newton-Kantorovich error bound $\frac{1}{2} q \lambda_{\min}$ (see Theorem 3.2) below $10^{-8}$ (see right hand side of Figure 7). In the depicted setting, the surface is discretized by triangle meshes with 64309 vertices and the uniform measure $\nu_0$ on $S^1$ is discretized by $n = 720$ quadrature points. Our test machine performed the task in about 2.6 seconds.
Figure 8: Performance comparison between Abikoff-Ye iteration and the damped, regularized Newton method (DRNM). We applied both methods to $N$ randomly generated polygonal lines with $n$ equal length edges and recorded the average number of iterations and the average time per polygonal line. The shaded regions in the iteration plots indicate the region between the $p/2$- and $(100 - p/2)$-percentiles, where $p \in \{0, 10, 50\}$. Top: Unit edge vectors uniformly sampled over $S^2$. Center: Unit edge vectors sampled from the von Mises-Fisher distribution with $\kappa = 1$. Bottom: An example that illustrates that “almost closed” does not necessarily imply “easy to close”. DRNM required at most 5 iterations per polygon whereas Abikoff-Ye required almost 3000 iterations in some cases.
6.2 Polygon Closure

We next present the results of experiments on polygon closures. Recall from Section 1 that
Millson-Kapovich showed that the space of polygons with fixed edgelengths is a quotient of
the space of nice semi-stable point measures on the sphere by the action of the conformal
group. So we can close a polygonal line by computing the conformal barycenter $w_s(\mu)$ of
$\mu = \sum_{i=1}^n \omega_i \delta(x_i)$ and by applying the shift transformation $\sigma_{w_s(\mu)}$ to the unit edge vectors
$x_1, \ldots, x_n$. Here $\omega_1, \ldots, \omega_n$ are the edgelengths of the polygonal line, normalized to satisfy
$\omega_1 + \cdots + \omega_n = 1$.

Figure 8 shows performance results of Abikoff-Ye iteration and of the damped, regularized
Newton method (DRNM) that we analyzed here. Both algorithms are accessible from the
routine ConformalBarycenter in the aforementioned Mathematica package. They were
applied to polygonal lines whose unit edge vectors were randomly sampled over the unit sphere $S^2$: In the first example we sampled the unit edge vectors uniformly over $S^2$. In the second
elementary example, we sampled from the spherical von Mises-Fisher distribution
$\varrho(x) = C_\kappa \exp(\kappa \langle \xi, x \rangle)$ with $\kappa = 1$. Here $\xi \in S^2$ is an arbitrary unit vector and $C_\kappa$ is a normalizing constant such that
$\int_{S^2} \varrho(x) \, dH^2(x) = 1$. For this distribution, the unit edge vectors are slightly concentrated
around $\xi$, resulting in “straighter” polygonal lines. Surprisingly, this does not substantially
increase the difficulty of the closing problem. We see that both iteration count and runtime
of the Abikoff-Ye iteration grow linearly in terms of the desired “accuracy” $\log(1/\varepsilon)$, where
we used (37) as stopping criterion. This is expected as the Abikoff-Ye iteration boils down to
steepest descent. In contrast, DRNM performs like Newton’s method (because we disabled line
search as soon as $q_k < 1$) with iteration count and runtime depending sublinearly on $\log(1/\varepsilon)$. It can also be seen from Figure 8 that the spread of iteration counts for DRNM is substantially
smaller than for Abikoff-Ye iteration.6

The last example does probably not reflect the typical use case, but it points directly onto
the weak spot of the Abikoff-Ye iteration. Here we sampled from a much more concentrated
spherical von Mises-Fisher distribution ($\kappa = 20$), but we also introduced a sharp kink of about
180° after half the number of edges. This produces “almost closed” polygonal lines (i.e.,
$w_s(\mu) \approx 0$) whose unit edge vectors are concentrated around two distinct points on the sphere.
As we saw in Proposition 5.3, this leads to a very high condition number of the Hessian of
$\Psi_\mu$. It is well-known that this case is particularly bad for steepest descent if this happens at the
minimizer $w_s(\mu)$. We can see the problem illustrated in the third row of Figure 8. In contrast,
DRNM is largely unaffected by this severe loss of conditioning; it converges in no more than
five iterations.

7 Conclusion and Future Directions

We have now given two algorithms for computing the conformal barycenter: Newton’s method
with fixed stepsize (16) and the regularized Newton method with line search (29)–(30). The

4See https://github.com/HenrikSchumacher/ConformalBarycenter for a maintained version.
5Intel Core i7 4980HQ CPU (2.8 GHz Quad-Core) with 16 GB RAM.
6The per-polygon timings of both methods are just too small to be timed accurately, so we refrained from
determining their spread.
first algorithm is primarily interesting as a device for proving theorems: Corollary 3.3 and Theorem 4.1 show that in all but exponentially few cases, the conformal barycenter can be approximated to fixed accuracy in linear time. In practical implementations, we recommend the use of (34)–(36), which converges for all input cases with a solution (Theorem 5.9) and in practice does so even faster than (16). We note that checking the Newton-Kantorovich condition of Corollary 3.3 reduces to computing the smallest eigenvalue of a $d \times d$ matrix, so one can switch between algorithms if a hard bound is desirable.

We think that it is probably possible to establish an explicit time bound for our second algorithm in terms of the parameters $\epsilon$ and $\delta$ from Lemma 5.6 and Lemma 5.8. However, we have deferred this problem until a compelling reason to solve it arises.

Our original interest in the conformal barycenter was motivated by its presence in the polygon and arclength parametrized closed curve constructions of Millson and coauthors [17, 3]. We intend to follow up on making this construction effective in computational geometry using the methods presented above. The Douady-Earle surfaces above were an intriguing surprise. What geometric properties do they have? This seems an avenue worth more investigation.

Acknowledgments

This work was supported by a postdoc fellowship of the German Academic Exchange Service (DAAD). We are also grateful for the support of the Simons Foundation (#524120 to Cantarella). In addition, many colleagues and friends contributed insightful discussions about the conformal barycenter, including Kyle Chapman, Philipp Reiter, Erik Schreyer, and Clayton Shonkwiler.

References

[1] Adrien Douady and Clifford J Earle. Conformally natural extension of homeomorphisms of the circle. Acta Mathematica, 157(1-2):23–48, 1986.

[2] Zvi Drezner, Kathrin Klamroth, Anita Schöbel, and George O Wesolowsky. The Weber Problem. In Horst W Hamacher and Zvi Drezner, editors, Facility Location, pages 1–36. Springer-Verlag, Berlin, 2002.

[3] Michael Kapovich and John J. Millson. The symplectic geometry of polygons in Euclidean space. J. Differential Geom., 44(3):479–513, 1996.

[4] Erik D. Demaine and Joseph O’Rourke. Geometric folding algorithms. Cambridge University Press, Cambridge, 2007. Linkages, origami, polyhedral.

[5] William Abikoff and Taiping Ye. Computing the Douady-Earle extension. In Lipa’s legacy (New York, 1995), volume 211 of Contemp. Math., pages 1–8. Amer. Math. Soc., Providence, RI, 1997.

[6] William Abikoff. Conformal barycenters and the Douady-Earle extension—a discrete dynamical approach. J. Anal. Math., 86:221–234, 2002.
[7] O. P. Ferreira and B. F. Svaiter. Kantorovich’s theorem on Newton’s method in Riemannian manifolds. *J. Complexity*, 18(1):304–329, 2002.

[8] Jason Cantarella, Kyle Chapman, Philipp Reiter, and Clayton Shonkwiler. Open and closed random walks with fixed edgelengths in $\mathbb{R}^d$. *J. Phys. A*, 51(43):434002–25, September 2018.

[9] Wolfgang Ring and Benedikt Wirth. Optimization methods on Riemannian manifolds and their application to shape space. *SIAM J. Optim.*, 22(2):596–627, 2012.

[10] Serge Lang. *Differential and Riemannian manifolds*, volume 160 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, third edition, 1995.

[11] Hans Schwerdtfeger. *Geometry of complex numbers*. Dover Publications, Inc., New York, 1979. Circle geometry, Moebius transformation, non-Euclidean geometry, A corrected reprinting of the 1962 edition, Dover Books on Advanced Mathematics.

[12] William M. Goldman. *Complex hyperbolic geometry*. Oxford Mathematical Monographs. The Clarendon Press, Oxford University Press, New York, 1999. Oxford Science Publications.

[13] Manfredo Perdigão do Carmo. *Riemannian geometry*. Mathematics: Theory & Applications. Birkhäuser Boston Inc., Boston, MA, 1992. Translated from the second Portuguese edition by Francis Flaherty.

[14] C. J. Atkin. The Hopf-Rinow theorem is false in infinite dimensions. *Bull. London Math. Soc.*, 7(3):261–266, 1975.

[15] Roman A. Polyak. Complexity of the regularized Newton’s method. *Pure Appl. Funct. Anal.*, 3(2):327–347, 2018.

[16] Jorge Nocedal and Stephen J. Wright. *Numerical optimization*. Springer Series in Operations Research and Financial Engineering. Springer, New York, second edition, 2006.

[17] John J Millson and Brett Zombro. A Kähler structure on the moduli space of isometric maps of a circle into euclidean space. *Inventiones Mathematicae*, 123(1):35–59, 1996.