Computing the Fréchet distance between real-valued surfaces

Computing the Fréchet Distance between Real-Valued Surfaces

Kevin Buchin[*] Tim Ophelders[*] Bettina Speckmann[*]

Abstract

The Fréchet distance is a well-studied measure for the similarity of shapes. While efficient algorithms for computing the Fréchet distance between curves exist, there are only few results on the Fréchet distance between surfaces. Recent work has shown that the Fréchet distance is computable between piecewise linear functions f and $g \colon M \to \mathbb{R}^k$ with M a triangulated surface of genus zero. We focus on the case $k = 1$ and M being a topological sphere or disk with constant boundary. Intuitively, we measure the distance between terrains based solely on the height function. Our main result is that in this case computing the Fréchet distance between f and g is in NP.

We additionally show that already for $k = 1$, computing a factor $2 - \varepsilon$ approximation of the Fréchet distance is NP-hard, showing that this problem is in fact NP-complete. We also define an intermediate distance, between contour trees, which we also show to be NP-complete to compute. Finally, we discuss how our and other distance measures between contour trees relate to each other.

Acknowledgements. K. Buchin, T. Ophelders, and B. Speckmann are supported by the Netherlands Organisation for Scientific Research (NWO) under project no. 612.001.207 (K. Buchin) and no. 639.023.208 (T. Ophelders & B. Speckmann).

[*]TU Eindhoven, The Netherlands. `[k.a.buchin|t.a.e.ophelders|b.speckmann]@tue.nl`

1 Introduction

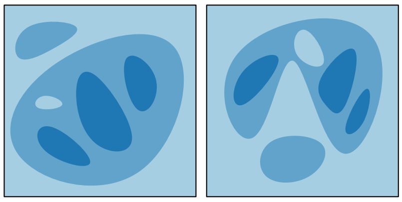

Figure 1: Two pictures that can continuously deform into each other.

The problem of measuring the similarity between shapes has recently gained much attention. While many measures have been defined, algorithms to compute such measures have been found for only a few of them. We consider the problem of comparing real-valued functions $f\colon M \to \mathbb{R}$ on surfaces, focusing in particular on spheres and disks of constant boundary, i.e., $f(x) = f(x')$ for all $x, x' \in \partial M$. The kind of similarity we investigate is that under continuous deformations of surfaces, such as in Figure 1. Here shapes that can be deformed into each other have distance 0, otherwise, shapes have some meaningful positive distance. Two natural computational problems arise for each measure, namely deciding whether two images have distance 0, and the more general problem of computing the distance between two images.

Major applications of computing such measures are in the comparison of medical imagery. For example, when comparing two MRI or CT scans of lungs, the images are often not aligned due to breathing and gravity. It is important to align the images through deformation to locate differences. A problem with similar use cases is elastic image matching, which is known to be NP-complete [27]. In practice, this problem is currently approached using various heuristics [34].

Definitions, background and results. Given two functions $f\colon M \to \mathbb{R}^k$ and $g\colon M \to \mathbb{R}^k$ with common parameter space M, their *Fréchet distance* [23, 24] is defined as

$$d_F(f,g) = \inf_{\mu\colon M \to M} \sup_{x\colon M} d(f(x), g \circ \mu(x)),$$

where $\mu\colon M \to M$ ranges over orientation preserving homeomorphisms and $d(\cdot,\cdot)$ is the underlying norm of $\mathbb{R}^k$. Essentially, the Fréchet distance captures the similarity between two functions by realigning their parameter spaces to minimize the maximum difference in function value of aligned points. It is often assumed that f and g are piecewise-linear functions.

Efficient algorithms for computing $d_F(f,g)$ exist for L^p norms if f and g are polylines [5], so if $M = [0,1]$ or $M = S^1$ for closed polylines. The computational complexity of this case is well understood with a significant amount of recent work addressing the fine-grained complexity of this case and its variants [1, 10, 13, 15]. Considerable progress has also been made recently on the problem of approximating the Fréchet distance between polylines [11, 21]. With many efficient algorithms at hand the Fréchet distance between polylines has found numerous applications, with recent work particularly focusing on geographic applications such as map-matching tracking data [9] and moving objects analysis [12].

The computational complexity in the case that f and $g\colon M \to \mathbb{R}^k$ are (triangulated) surfaces is much less understood. The problem is known to be NP-hard [26] also when $k = 2$ [14, 17]. But it is not known whether it is actually in NP, in fact until recently it was only known to be upper semi-computable for surfaces in $\mathbb{R}^k$ [4, 18]. Recently, the first computability result [30] for surfaces of genus 0 was found, including an exponential time approximation algorithm.

We show that for $k = 1$, the problem is in NP if M is a topological sphere or disk with constant boundary. Additionally, we show that already for $k = 1$ computing the Fréchet distance is NP-hard. Even for this case, no efficient constant-factor approximation algorithms have been found so far, and we show that computing a factor $2 - \varepsilon$ approximation is NP-hard.

In previous work, a few variants on the comparison of surfaces under the Fréchet distance have been investigated. For instance, there are efficient algorithms for computing the Fréchet distance with certain constraints on the homeomorphisms μ [17] and for computing the *weak Fréchet distance* [4] between triangulated surfaces homeomorphic to the disk. The Fréchet distance can be computed in polynomial time for simple polygons (including interior) in the plane [16], even if they have one [14] or more holes [29]. An efficient constant factor approximation exists for so-called folded polygons, which need not lie in the plane [19].

In addition to our results on surfaces (Section 3), we define a suitable similarity measure between contour trees, which we show to be NP-complete to compute as well (Section 2). Recently, other measures [6, 7, 8, 20, 28] such as the interleaving distance or functional distortion distance have been investigated for contour trees. In particular, it is expected that the interleaving distance and the functional distortion distance are equal, but so far it has been proven only that they differ by at most a constant factor. For computing the interleaving distance, no efficient exact algorithms are known. We also show a relation to the Gromov-Hausdorff distance between trees, for which an NP-hardness proof was recently found [2]. The same paper proposes polynomial-time approximation algorithms with an approximation ratio of $O(\min(n, \sqrt{rn}))$ (where r is the length ratio between the longest and shortest edge) for the Gromov-Hausdorff and interleaving distances. We highlight how the Fréchet distance and weak Fréchet distance differ from these measures in Section 5.

Outline. Our main result is the NP-completeness of the Fréchet distance between $\mathbb{R}$-valued surfaces. Because of its combinatorial complexity, we start by defining a combinatorially simpler problem, the contour tree distance. The contour tree distance is a similarity measure related to the Fréchet distance, that abstracts from the geometric representation of the matching μ. As such, the contour tree distance is a lower bound on the Fréchet distance.

We show that the contour tree distance is in NP using properties of the freespace diagram, which is a commonly used tool in algorithms for computing the Fréchet distance between curves. Using extensions of the freespace diagram, valid matchings between contour trees can be analyzed easily. Nonetheless, it turns out that approximating the contour tree distance within a factor 2 is NP-hard by a reduction from the NP-hard problem Exact Cover by 3-sets.

Using specific properties of the matchings used in this reduction, these matchings can be extended to matchings μ between surfaces. A reduction from the contour tree distance to the Fréchet distance between $\mathbb{R}$-valued surfaces then shows that approximating the Fréchet distance within a factor 2 is also NP-hard. However, not all matchings between contour trees have an extension to a matching between surfaces, as discussed in Section 4.

To show that computing the Fréchet distance is in NP, we make use of a polynomial amount of information of a matching μ. Using this information, we can represent a matching as an Euler diagram, which can be recognized in NP.

2 Contour tree distance

The Reeb graph [31] of a function $f\colon M \to \mathbb{R}$ is the quotient space $M/{\sim_f}$ where $a \sim_f b$ if and only if a and b are in the same connected component of the level set $f^{-1}(f(a))$. Denote by $\mathcal{R}_f$ the corresponding quotient map. Because f associates a single real number to each equivalence class of $\sim_f$, the resulting Reeb graph has a natural $\mathbb{R}$-valued function associated with it, namely the (unique) function $f'\colon M/{\sim_f} \to \mathbb{R}$ satisfying $f' \circ \mathcal{R}_f = f$. If M is the disk or the 2-sphere, the Reeb graph forms a tree called a contour tree.

For the sake of compact representation, in this paper we assume each surface to be triangulated, to form a simplicial 2-complex. Moreover, we assume without loss of generality that function values along edges of Reeb graphs are interpolated linearly between the values of the vertices at their endpoints. In this representation, the contour tree of a surface with n faces has complexity $O(n)$ and can be computed in $O(n \log n)$ time [35]. Intuitively, the contour tree has a vertex for each minimum, maximum and saddle point, and edges between 'adjacent' vertices.

Based on the Fréchet distance between f and g, we derive a computationally simpler measure that abstracts from the realizability of the matching μ between spheres or disks. Throughout this paper, we use the notation $\mathbb{X} = M/{\sim_f}$ and $\mathbb{Y} = M/{\sim_g}$ for the contour trees of f and g, respectively. We shall denote the vertex set of $\mathbb{X}$ by $V(\mathbb{X})$ and its edge set by $E(\mathbb{X})$. With slight abuse of notation, we reuse function names f and g for the natural $\mathbb{R}$-valued functions associated with the contour trees $\mathbb{X}$ and $\mathbb{Y}$. Our distance measure d_C compares the contour trees $\mathbb{X}$ and $\mathbb{Y}$ of f and g. We define the contour tree distance d_C as

$$d_C(f\colon \mathbb{X} \to \mathbb{R}, g\colon \mathbb{Y} \to \mathbb{R}) = \inf_{\tau \in \mathcal{M}(\mathbb{X},\mathbb{Y})} \sup_{(x,y)\in\tau} |f(x) - g(y)|,$$

where $\tau \subseteq \mathbb{X} \times \mathbb{Y}$ is drawn from some class of matchings $\mathcal{M}(\mathbb{X},\mathbb{Y})$, defined below. So τ defines a correspondence between contour trees, such that $(x, y) \in \tau$ if some points on contours x and y were matched by a corresponding matching μ on M. Denote $\tau(x) = \{y \mid (x, y) \in \tau\}$ and $\tau^{-1}(y) = \{x \mid (x, y) \in \tau\}$. The class $\mathcal{M}(\mathbb{X},\mathbb{Y})$ captures the essential (but not all) properties of an orientation preserving matching μ. We define $\mathcal{M}(\mathbb{X},\mathbb{Y})$ as the set of matchings τ for which the following properties hold:

1. τ is a connected subset of $\mathbb{X} \times \mathbb{Y}$;
2. $\tau(x)$ is a nonempty subtree of $\mathbb{Y}$ for each $x \in \mathbb{X}$;
3. $\tau^{-1}(y)$ is a nonempty subtree of $\mathbb{X}$ for each $y \in \mathbb{Y}$.

Here, the term *subtree* is used to denote a connected subset of a tree, not necessarily containing leaves of that tree. By Conditions 2 and 3, each connected set matches with a connected set, and Condition 1 ensures continuity. Let $x \rightsquigarrow_{\mathbb{X}} x'$ denote all (possibly backtracking) paths from x to x' in the underlying space $\mathbb{X}$. The following properties can be derived immediately:

1. $\tau(\mathbb{X}) = \mathbb{Y}$ and $\tau^{-1}(\mathbb{Y}) = \mathbb{X}$;
2. for each $p_f\colon x \rightsquigarrow_{\mathbb{X}} x'$, some path $p\colon (x, y) \rightsquigarrow_{\mathbb{X}\times\mathbb{Y}} (x', y')$, whose projection onto $\mathbb{X}$ equals p_f, satisfies $p \subseteq \tau$;
3. for each $p_g\colon y \rightsquigarrow_{\mathbb{Y}} y'$, some path $p\colon (x, y) \rightsquigarrow_{\mathbb{X}\times\mathbb{Y}} (x', y')$, whose projection onto $\mathbb{Y}$ equals p_g, satisfies $p \subseteq \tau$;
4. if $\{(x, y), (x', y)\} \subseteq \tau$ then $p \times \{y\} \subseteq \tau$ for some $p\colon x \rightsquigarrow_{\mathbb{X}} x'$;
5. if $\{(x, y), (x, y')\} \subseteq \tau$ then $\{x\} \times p \subseteq \tau$ for some $p\colon y \rightsquigarrow_{\mathbb{Y}} y'$.

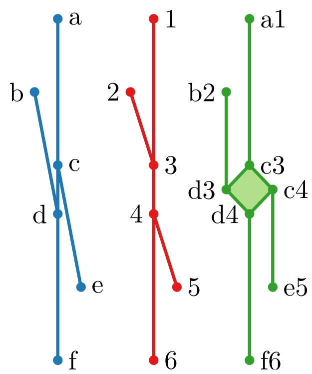

Figure 2: Two trees (left) and a matching (right).

An example of a matching between two trees is shown in Figure 2. The two-dimensional patch illustrates a many-to-many correspondence. For a matching $\mu\colon M \to M$ between surfaces f and g, define τ_μ to be the corresponding matching between the Reeb graphs of f and g:

$$\tau_\mu = \{(\mathcal{R}_f(x), \mathcal{R}_g \circ \mu(x)) \mid x \in M\}.$$

By Lemma 1 we have for each orientation preserving homeomorphism μ, that the matching $\tau_\mu \in \mathcal{M}(\mathbb{X},\mathbb{Y})$, and hence that $d_C(f, g) \le d_F(f, g)$. On the other hand, a matching $\tau \in \mathcal{M}(\mathbb{X},\mathbb{Y})$ does not need to correspond to an orientation preserving homeomorphism on M. We give an example of such a matching in Section 4.

Lemma 1 *$\tau_\mu \in \mathcal{M}(\mathbb{X},\mathbb{Y})$ for any orientation preserving homeomorphism $\mu\colon M \to M$.*

Proof. Consider such a matching μ. We show all three conditions of $\mathcal{M}(\mathbb{X},\mathbb{Y})$ for τ_μ. The set $\{(x,y) \mid \mu(x)=y\}$ is a connected subset of $S^2 \times S^2$. Hence its image under the quotient map $(x,y) \mapsto (\mathcal{R}_f(x), \mathcal{R}_g(y))$ is connected, so Condition 1 holds. Because μ is a homeomorphism on a connected set, $\tau_\mu(x)$ is connected, and by surjectivity of τ_μ^{-1} nonempty. So, $\tau_\mu(x) \subseteq \mathbb{Y}$ and symmetrically $\tau_\mu^{-1}(y) \subseteq \mathbb{X}$ is a nonempty subtree. $\square$

Corollary 1 $d_C(f,g) \leq d_F(f,g)$.

To test whether the contour tree distance between two trees is zero, one only needs to test whether the trees are equal. We represent trees canonically by exhaustively removing degree 2 vertices that lie on the line segment connecting the two adjacent vertices, and replacing them by a single edge between those vertices. This reduces the problem of testing a contour tree distance of zero to labeled unordered unrooted tree isomorphism, which can be solved in linear time [3].

It should be noted that the contour tree distance and Fréchet distance between trees are different problems. In fact, one major limitation of the Fréchet distance for trees is that non-homeomorphic trees have infinite Fréchet distance. Nonetheless, an $O(n^{5/2})$ time algorithm [14] for computing the Fréchet distance between trees has been found.

2.1 Free space diagrams

We consider the decision problem that asks whether the contour tree distance between two $\mathbb{R}$-valued trees is at most ε. The *free space diagram* is a commonly used data structure in algorithms for computing the Fréchet distance. We can also formulate a *freespace* in relation to the contour tree distance. Define the *ε-freespace* $\mathcal{F}_\varepsilon$ of f and g to be the set of pairs in the product of the parameter spaces whose respective images are at most ε apart:

$$\mathcal{F}_\varepsilon = \{(x,y) \in \mathbb{X} \times \mathbb{Y} \mid |f(x) - g(y)| \leq \varepsilon\}.$$

Define an *ε-matching* to be a matching $\tau \in \mathcal{M}(\mathbb{X},\mathbb{Y})$ for which all $(x,y) \in \tau$ satisfy $|f(x)-g(y)| \leq \varepsilon$. Then an ε-matching $\tau \subseteq \mathcal{F}_\varepsilon$ exists if (and only if) the answer to the decision problem is yes. The contour tree distance is the minimum value of ε for which an ε-matching exists. In Figure 4, a product parameter space containing the ε-matching of Figure 3 is drawn in green as a projection from four-dimensional space.

Without loss of generality, assume function values along edges of $\mathbb{X}$ and $\mathbb{Y}$ to be interpolated linearly. A commonly used property of the freespace is that for each edge $e_x \in E(\mathbb{X})$ and edge $e_y \in E(\mathbb{Y})$, the face $e_x \times e_y \in E(\mathbb{X}) \times E(\mathbb{Y})$ has a convex intersection with the freespace $\mathcal{F}_\varepsilon$ (see Lemma 2, originally proven in [5]). An important consequence of this lemma which we will see in Section 2.2 is that given only information about the intersection of the ε-freespace with the four edges bounding such a face, sufficient information about the interior of the face can be derived to decide whether an ε-matching exists.

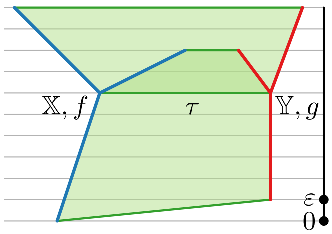

Figure 3: Trees $\mathbb{X}$ and $\mathbb{Y}$ with an ε-matching τ.

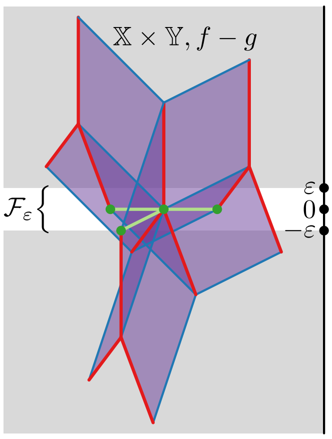

Figure 4: Matching τ in the freespace $\mathcal{F}_\varepsilon$.

Lemma 2 *For an edge $e_x \in E(\mathbb{X})$ and an edge $e_y \in E(\mathbb{Y})$ on which the respective functions f and g are linear, the face $F = e_x \times e_y$ has a convex intersection with $\mathcal{F}_\varepsilon$ for any ε.*

Proof. The map $(x,y) \mapsto f(x) - g(y)$ is affine when restricted to domain F. The preimage of a convex set (the ε-ball) under an affine map is convex. Intersecting this convex preimage with the convex set F yields again a convex set. Hence $F \cap \mathcal{F}_\varepsilon$ is convex. $\square$

Let ε be the value of the contour tree distance, then ε is the minimum value for which all constraints of $\mathcal{M}(\mathbb{X}, \mathbb{Y})$ can be satisfied by some matching τ. This means that ε is the minimum value for which either some vertex $(x, y) \in V(\mathbb{X}) \times V(\mathbb{Y})$ lies in $\mathcal{F}_\varepsilon$, or for two (not necessarily distinct) edges $e_1, e_2 \in V(\mathbb{X}) \times E(\mathbb{Y})$ or $e_1, e_2 \in E(\mathbb{X}) \times V(\mathbb{Y})$ of $\mathbb{X} \times \mathbb{Y}$, the sets $e_1 \cap \mathcal{F}_\varepsilon$ and $e_2 \cap \mathcal{F}_\varepsilon$ contain points with the same y- or x-coordinates, respectively. In case ε is such minimum value, we call ε a *critical value*, and there are only $O(n^3)$ of them. For two functions f and g the $O(n^3)$ critical values of ε can each be computed in constant time. However, determining which of those critical values determines the contour tree distance is NP-hard as shown in Section 2.3. First we show that computing the contour tree distance is in NP.

2.2 Contour tree distance in NP

We show that given only a polynomial amount of information about an ε-matching τ, it can be verified in polynomial time that an ε-matching $\tau' \in \mathcal{M}(\mathbb{X}, \mathbb{Y})$ exists. The information we use is for each vertex x of $\mathbb{X}$ the endpoints and vertices of the corresponding subtree $\tau(x)$ of $\mathbb{Y}$, and for every vertex y of $\mathbb{Y}$ the endpoints and vertices of the corresponding subtree $\tau^{-1}(y)$ of $\mathbb{X}$. This set of points P_τ is defined more formally in Equation 1. To illustrate, if vertex x maps to the subtree $\tau(x)$ of $\mathbb{Y}$ shown in Figure 5, then exactly the four bullets of $\mathbb{Y}$ drawn in green will appear paired with x in P_τ. For any matching $\tau \in \mathcal{M}(\mathbb{X}, \mathbb{Y})$, there are only $O(n^2)$ points in the set P_τ. However, it should be noted that such points give no information about the behavior of τ in the interior of faces of $\mathbb{X} \times \mathbb{Y}$.

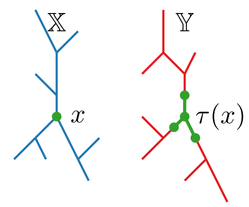

Figure 5: A subtree $\tau(x)$ of $\mathbb{Y}$ (green) matched with a vertex x of $\mathbb{X}$.

$$P_\tau = \{(x, y) \mid x \in V(\mathbb{X}) \wedge y \in \partial(\tau(x))\} \;\cup\; \{(x, y) \mid y \in V(\mathbb{Y}) \wedge x \in \partial(\tau^{-1}(y))\} \cup (\tau \cap (V(\mathbb{X}) \times V(\mathbb{Y}))), \text{ where } \partial \text{ is the boundary operator.} \tag{1}$$

Let $CH(P)$ be the convex hull (including its interior) of a point set P. For a given set P_τ, derive τ' using Equation 2; such that for each face F, τ' consists of the convex hull of $F \cap P_\tau$. In Lemma 3, we show that τ' is a valid ε-matching if P_τ is derived from an ε-matching τ.

$$\tau' = \bigcup_{F \in E(\mathbb{X}) \times E(\mathbb{Y})} CH(F \cap P_\tau). \tag{2}$$

Lemma 3 *If $\tau \in \mathcal{M}(\mathbb{X}, \mathbb{Y})$ is an ε-matching, then τ' derived from P_τ is an ε-matching with $\tau' \in \mathcal{M}(\mathbb{X}, \mathbb{Y})$.*

Proof. If τ was an ε-matching, then by Lemma 2, τ' is a subset of $\mathcal{F}_\varepsilon$. It remains to verify the three conditions of $\mathcal{M}(\mathbb{X}, \mathbb{Y})$ for τ'. Any face of $\mathbb{X} \times \mathbb{Y}$ will contain at most one component of τ by convexity. We show that all nonempty faces are connected, indeed the intersection of the boundary of a face with τ' is the same as that with τ. Hence τ' cannot have more components than τ. Then because τ is connected, τ' must also be connected, satisfying Condition 1.

For each vertex $x \in V(\mathbb{X})$, we know that $\tau'(x)$ is a nonempty tree. We show that this holds even if x is internal to some edge $e_x \in E(\mathbb{X})$. Because τ' is connected and has nonempty trees at the endpoints of e_x, the set $\tau'(x)$ cannot be empty. So suppose for contradiction that $\tau'(x)$ has multiple components, then because the interior of each face is convex, we must have $y \notin \tau'(x)$ for some vertex $y \in \tau(x)$ which is not possible. So Condition 2 and by a symmetric argument Condition 3 holds. $\square$

Therefore, if ε is the contour tree distance, a set P_τ of polynomial size exists, whose derived τ' is an ε-matching. Given P_τ, it is easily tested whether for its derived τ', the properties required by $\mathcal{M}(\mathbb{X}, \mathbb{Y})$ are satisfied. Moreover, as a consequence of Lemma 2, it can be verified that τ' is an ε-matching by checking that $|f(x) - g(y)| \leq \varepsilon$ for all $(x, y) \in P_\tau$. Hence, Theorem 1 follows.

Theorem 1 *Deciding whether the contour tree distance between $\mathbb{X}$ and $\mathbb{Y}$ is at most ε is in NP.*

Remark. The proof that the contour tree distance is in NP extends to trees $\mathbb{X}$ and $\mathbb{Y}$ in higher dimensions. Specifically, if $f\colon \mathbb{X} \to \mathbb{R}^k$ and $g\colon \mathbb{Y} \to \mathbb{R}^k$, then computing $d_C(f, g) = \inf_{\tau \in \mathcal{M}(\mathbb{X},\mathbb{Y})} \sup_{(x,y)\in\tau} \|f(x) - g(y)\|_p$ is also in NP for L^p norms $\|\cdot\|_p$.

2.3 NP-hardness

We show that approximating the contour tree distance between $\mathbb{R}$-valued trees within factor 2 is NP-hard by a reduction from the NP-hard problem Exact Cover by 3-sets [25].

Definition 1 Exact Cover by 3-sets *(X3C)*
Input: A set $\mathcal{S}$ of m subsets of $\{1, \ldots, k\}$, each of size 3.
Output: Does a subset (consisting of $k/3$ triples) of $\mathcal{S}$ partition $\{1, \ldots, k\}$?

We introduce the gadgets used in our reduction in Figure 6. For this, define a zig-zag of radius r centered at position p along a segment $[a, b]$ to be a path visiting vertices at positions a, $p+r$, $p-r$ and b, in that order. Gadget Y^* is a long segment from position 0 to position $6k + 6$. Gadget Y_l (with $l \in \{1, \ldots, k\}$) is a path from position 1 to $6k+6$ with a zig-zag of radius 2 around position $6l$. Similarly, gadget $X_{i,j}$ ($i \in \{1, \ldots, m\}$, $j \in \{1, 2, 3\}$) has a zig-zag around position $6 \cdot s(i, j)$, but with radius 1. The function s aligns the center of the zig-zag of $X_{i,j}$ with that of $Y_{s(i,j)}$, such that gadget $X_{i,j}$ has a contour tree distance of 1 to Y^* and $Y_{s(i,j)}$, but a contour tree distance of 2 to any gadget Y_l with $l \neq s(i, j)$.

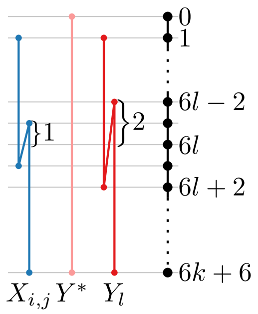

Figure 6: Gadgets.

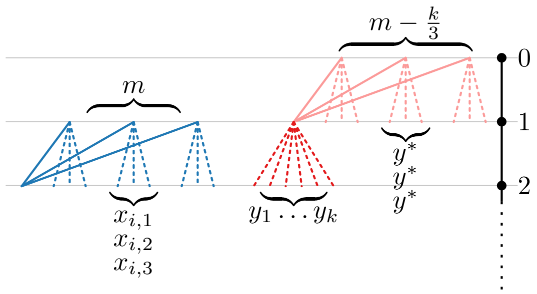

Figure 7: Gadgets are connected to form trees f (blue) and g (red).

The function s can be configured such that each triple of gadgets $(X_{i,1},\ X_{i,2},\ X_{i,3})$ corresponds to one of the m subsets of $\mathcal{S}$. We connect the three elements of each triple at a common vertex at position 1, and finally connect all triples at a common vertex at position 2 (blue in Figure 7) to form tree $f\colon \mathbb{X} \to \mathbb{R}$.

Similarly, each gadget Y_l corresponds to an element of $\{1, \ldots, k\}$, and all Y_l are connected to a common vertex at position 1. The idea is that to obtain a low contour tree distance, $k/3$ triples of f must match the Y_l gadgets exactly; then what remains in f are $m - k/3$ triples that must be matched elsewhere. Each such unmatched triple of f will then be forced to match with three copies of Y^*, connected at a vertex at position 0 to form a so called Y^*-triple. We use $m - k/3$ such Y^*-triples, each connected to the Y_l gadgets at position 1 to form tree $g\colon \mathbb{Y} \to \mathbb{R}$. In Lemma 4, we use a solution to X3C to derive a matching using only many-to-one correspondences between f and g, even though many-to-many correspondences are permitted by $\mathcal{M}(\mathbb{X}, \mathbb{Y})$.

Lemma 4 *$d_C(f,g) \leq 1$ if $\mathcal{S}$ admits a solution to* X3C*, even if only many-to-one correspondences are allowed.*

Proof. Consider the correspondence (as given by a solution to X3C) between gadgets $X_{i,j}$ of f and gadgets Y_l of g, and let each Y^*-triple correspond to an unused $(X_{i,1},\ X_{i,2},\ X_{i,3})$-triple of f. Then each individual gadget has distance at most 1 to the corresponding gadget of the other tree. The connections between gadgets (see Figure 7) also have a correspondence of distance at most 1 by matching connections of f that are part of the solution with the vertex of g at position 1 and matching remaining connections with connections to Y^*-triples. □

Lemma 5 *$d_C(f,g) \geq 2$ unless $\mathcal{S}$ admits a solution to* X3C*.*

Proof. Assume instead that $d_F(f,g) < 2$, then for some matching τ between f and g, the distance between all matched points is less than 2.

The endpoint $\textsc{End}(x_{i,j})$ at position $6k+6$ for each gadget $X_{i,j}$ of f must be matched to a connected component of the preimage of $(6k+4, 6k+8)$ under g. Since gadgets are connected only in the interval $[0,2]$ which is disjoint from $(6k+4, 6k+8)$, each such component corresponds to a single gadget of g. Therefore each $X_{i,j}$ is matched with a single gadget of g. By a symmetric argument, the endpoint of each gadget of g is matched with a single gadget of f. Hence, a one-to-one correspondence between gadgets of f and g exists.

Now suppose two gadgets $X = X_{i,j}$ and $X' = X_{i',j'}$ of different triples (i.e. $i \neq i'$) correspond to two gadgets of the same Y^*-triple. Then each path connecting X and X' contains a point at position 2 which must be matched with the vertex at 0 that connects the two Y^* gadgets, yielding a contradiction. Thus, each triple $(X_{i,1},\ X_{i,2},\ X_{i,3})$ corresponds either to a single Y^*-triple, or three Y_l gadgets. Suppose $X_{i,j}$ corresponds to Y_l with $l \neq s(i,j)$, then the zig-zag of Y_l corresponds to a monotone path (without a zig-zag) of $X_{i,j}$. Let the vertices of Y_l at positions $6l+2$ and $6l-2$ be denoted v^+ and v^- respectively. By continuity of τ, we have $6l < f(\tau^{-1}(v^+)) \leq f(\tau^{-1}(v^-)) < 6l$, which is impossible. Hence, the contour tree distance between f and g is at least 2 if $\mathcal{S}$ admits no solution to X3C. □

Combining these lemmas with Theorem 1 we obtain Theorem 2.

Theorem 2 *Computing a $(2-\varepsilon)$-approximation of the contour tree distance is NP-complete, even if only many-to-one correspondences are allowed.*

3 Surfaces

We wish to use the fact that computing the contour tree distance is in NP to prove that the Fréchet distance between $\mathbb{R}$-valued surfaces is also in NP. We will consider two surfaces: the disk $[0,1]^2$ and the sphere S^2. It turns out that not all matchings between the contour trees $\mathbb{X}$ and $\mathbb{Y}$ can be realized as orientation preserving homeomorphisms on the sphere. An example of such a matching is described in Section 4. In the case of the disk, the boundaries must also be matched, which imposes additional constraints on the matching of the interiors.

We show in Section 3 that a polynomial amount of information about an ε-matching τ between contour trees is sufficient to decide in NP whether an ε-matching μ on the disk exists. We use this to prove that the Fréchet distance between $\mathbb{R}$-valued spheres or disks with constant boundary is in NP. For this we use properties of Euler diagrams, which are described in Section 3.1. The relation between matchings and Euler diagrams is discussed in Section 3.2, and how Euler diagrams can be used to derive matchings is discussed in Section 3.3.

3.1 Euler diagrams

An Euler diagram is a set of topological disks, drawn in the plane to capture relations such as overlap or containment between them. There are eight possible such relations [22], namely $Rel = \{$*disjoint*, *equal*, *inside*, *contains*, *covered*, *cover*, *meet*, *overlap*$\}$, whose meanings are as follows.

$$\begin{aligned}
\mathit{disjoint}(a,b) &\equiv a \cap b = \emptyset;\\
\mathit{equal}(a,b) &\equiv a = b;\\
\mathit{inside}(a,b) &\equiv a \subseteq b \text{ and } \partial a \cap \partial b = \emptyset;\\
\mathit{contains}(a,b) &\equiv \mathit{inside}(b,a);\\
\mathit{covered}(a,b) &\equiv a \subseteq b \text{ and } \partial a \cap \partial b \neq \emptyset;\\
\mathit{cover}(a,b) &\equiv \mathit{covered}(b,a);\\
\mathit{meet}(a,b) &\equiv a \cap b \subseteq \partial a \cap \partial b \neq \emptyset;\\
\mathit{overlap}(a,b) &\equiv a \not\subseteq b \text{ and } b \not\subseteq a.
\end{aligned}$$

For a set S of n elements, a *topological expression* [33, 36] on S is a function $\mathcal{P} : S \times S \to Rel$. We say a topological expression $\mathcal{P}$ on S is satisfiable if and only if there is an Euler diagram of disks $\{D(i) \mid i \in S\}$, such that for any $i, j \in S$, the relation between $D(i)$ and $D(j)$ is as given by $\mathcal{P}(i, j)$. It was shown by Schaefer, Sedgwick and Štefankovič [32] that deciding whether a topological expression $\mathcal{P}$ on n elements is satisfiable (by an Euler diagram in the plane) is in NP.

3.2 Euler diagrams from matchings

We show that deciding whether the Fréchet distance between $\mathbb{R}$-valued surfaces f and $g\colon M \to \mathbb{R}$ is at most ε is in NP if M is a sphere. First consider the case where $M = [0,1]^2$ is a topological disk and that the boundary of f as well as g has a constant function value, so $f(p) = f(q)$ and $g(p) = g(q)$ for all $p, q \in \partial M$.

Define a *refinement* of a contour tree $\mathbb{X}$ to be a tree homeomorphic to $\mathbb{X}$, whose edges are subdivided using extra vertices. Assume that $\mathbb{X}$ and $\mathbb{Y}$ both have a vertex (or add it otherwise) for the contour containing the boundary, which we treat as the root of the trees. We assume edges of rooted trees are directed away from the root.

For a vertex v of $\mathbb{X}$ with d outgoing edges, $\mathcal{R}_{\mathbb{X}}^{-1}(x)$ is a region homotopic to a d-holed disk on M. We index the boundary components of such a vertex by v and each of its outgoing edges, so that the set $W(\mathbb{X}) = V(\mathbb{X}) \cup E(\mathbb{X})$ of vertices and edges indexes all boundary components $\{c_w \mid w \in W(\mathbb{X})\}$ of vertices of $\mathbb{X}$. For a matching $\mu\colon M \to M$, for $w \in W(\mathbb{X})$ we are interested in the corresponding subtrees $\rho(w) = \mathcal{R}_{\mathbb{Y}} \circ \mathcal{R}_{\mathbb{X}}^{-1}(c_w)$ on $\mathbb{Y}$, and for $w \in W(\mathbb{Y})$ we are interested in the subtrees $\rho^{-1}(w) = \mathcal{R}_{\mathbb{X}} \circ \mu^{-1}(c_w)$ on $\mathbb{X}$. We can derive ρ given $\mu\colon M \to M$, and refine $\mathbb{X}$ into $\mathbb{X}'$ by adding vertices along edges at $\bigcup_{y \in W(\mathbb{Y})} \partial\rho^{-1}(y)$ and refine $\mathbb{Y}$ into $\mathbb{Y}'$ by adding vertices at $\bigcup_{x \in W(\mathbb{X})} \partial\rho(x)$. Since $\partial\rho^{-1}(y)$ and $\partial\rho(x)$ have $O(n)$ points per $y \in W(\mathbb{Y})$ or $x \in W(\mathbb{X})$, trees $\mathbb{X}'$ and $\mathbb{Y}'$ have $O(n^2)$ vertices.

For a vertex $x \in V(\mathbb{X}')$ of a rooted tree, let $sub(x) \subseteq \mathbb{X}$ be the subtree rooted at x, and for an edge $e = (x, x') \in E(\mathbb{X})$ from x to x', let $sub(e) = (e \cup sub(x')) \setminus \{x\}$. For each $i \in W(\mathbb{X}')$, there is a disk $D_\mu(i) = \mathcal{R}_f^{-1}(sub(i)) \subseteq M$, and for each $i \in W(\mathbb{Y}')$, there is a disk $D_\mu(i) = \mu^{-1}(\mathcal{R}_g^{-1}(sub(i))) \subseteq M$. We can now derive an Euler diagram from μ with $O(n^2)$ disks $\{D_\mu(i) \mid i \in W(\mathbb{X}') \cup W(\mathbb{Y}')\}$. The corresponding topological expression, call it $\mathcal{P}_\mu$, consists of $O(n^4)$ relations between these disks.

We capture the general structure of such expressions later, when we define topological expressions respecting τ_μ. For that, we first capture the structure of topological expressions for the disks of a single tree $\mathbb{X}'$, by saying a topological expression $\mathcal{P}$ on $S \supseteq W(\mathbb{X}')$ *respects* $\mathbb{X}$ if

1. for all $e = (i, j) \in E(\mathbb{X}')$, $\mathcal{P}(i, e) = \textit{contains}$ and $\mathcal{P}(e, j) = \textit{contains}$;
2. for all $e = (i, j) \in E(\mathbb{X}')$ and $e' = (i, j') \in E(\mathbb{X}')$ with $e \neq e'$, $\mathcal{P}(e, e') = \textit{disjoint}$.

For a satisfiable topological expression respecting $\mathbb{X}$, the drawing of the disks of $W(\mathbb{X}')$ is the same (up to orientation preserving homeomorphism) for any satisfying Euler diagram Φ. Let $\{D(i) \mid i \in S\}$ be the disks of such an Euler diagram Φ. Then, for $x \in V(\mathbb{X}')$, define ϕ_x as the closed region bounded by $\partial D(x)$ and $\bigcup_{e=(x,x')\in E(\mathbb{X}')} \partial D(e)$, which is a disk with a hole for each outgoing edge of x. For $e = (x, x') \in E(\mathbb{X}')$, define ϕ_e as the closed annulus bounded by $\partial D(e)$ and $\partial D(x')$. Denote $\phi_{V(\mathbb{X}')} = \bigcup_{x\in V(\mathbb{X}')} \phi_x$ and $\phi_{E(\mathbb{X}')} = \bigcup_{e\in E(\mathbb{X}')} \phi_e$.

We say a topological expression $\mathcal{P}$ on $W(\mathbb{X}') \cup W(\mathbb{Y}')$ respects a matching $\tau \in \mathcal{M}(\mathbb{X}, \mathbb{Y})$ if the following conditions hold.

1. $\mathcal{P}$ respects both $\mathbb{X}$ and $\mathbb{Y}$;
2. $\mathcal{P}(\textit{root}(\mathbb{X}'), \textit{root}(\mathbb{Y}')) = \textit{equal}$;
3. for $x \in V(\mathbb{X}')$ and $y \in V(\mathbb{Y}')$, ϕ_x intersects ϕ_y only if $y \in \tau(x)$ (and $x \in \tau^{-1}(y)$);
4. for each component of $\partial\phi_x$ with $x \in V(\mathbb{X}')$, it intersects ϕ_y for some $y \in V(\mathbb{Y}')$;
5. for each component of $\partial\phi_y$ with $y \in V(\mathbb{Y}')$, it intersects ϕ_x for some $x \in V(\mathbb{X}')$;

Any matching μ has an Euler diagram satisfying a topological expression respecting τ_μ.

Because we only use a polynomial amount of information about τ, and intersection relations between ϕ_x and ϕ_y can be formulated as topological expressions, it can be tested in NP that a topological expression respects τ, and by [32] that it is satisfiable.

3.3 Matchings from Euler diagrams

Call a matching μ an *ε-matching* if $|f(p) - g(\mu(p))| \leq \varepsilon$ for all $p \in M$. We show that there is some ε-matching μ for any satisfiable topological expression respecting some ε-matching τ. For this, let Φ be an Euler diagram satisfying a topological expression respecting ε-matching τ.

Let $M' = D(\textit{root}(\mathbb{X}')) = D(\textit{root}(\mathbb{Y}'))$. We construct functions f' and $g' : M' \to \mathbb{R}$ with $d_F(f', f) = 0$ and $d_F(g', g) = 0$, such that $\Delta(M') \leq \varepsilon$ where $\Delta(M') = \sup_{p\in M'} |f'(p) - g'(p)|$.

Start by assigning $f'(p) = f(x)$ for $p \in \phi_x$ for $x \in V(\mathbb{X}')$. Also assign $g'(p) = g(y)$ for $p \in \phi_y$ for $y \in V(\mathbb{Y}')$. Then $\Delta(\phi_{V(\mathbb{X}')} \cap \phi_{V(\mathbb{Y}')}) \leq \varepsilon$ because ϕ_x and ϕ_y intersect only if $f(x) - g(y) \leq \varepsilon$.

Next, we define g' on $\partial D(i)$ for each $i \in W(\mathbb{X}')$. By construction, $\partial D(i)$ intersects $\phi_{V(\mathbb{Y}')}$ for any $i \in W(\mathbb{X}')$, so g' is already defined for some point on $\partial D(i)$. We can hence assign g' to all points on $\partial D(i)$ by linear (arc length parameterized) interpolation between the points of $\phi_{V(\mathbb{Y}')} \cap \partial D(i)$. Because f' is constant on $\partial D(i)$, we have by interpolation that $\Delta(\partial D(i)) \leq \Delta(\phi_{V(\mathbb{Y}')} \cap \partial D(i)) \leq \varepsilon$. Assign f' analogously to $\partial D(i)$ for each $i \in W(\mathbb{Y}')$.

Let $\mathcal{F}$ be the faces of $M' \setminus \bigcup_{i\in W(\mathbb{X}')\cup W(\mathbb{Y}')} \partial D(i)$. Now, f' remains to be assigned for faces $F \in \mathcal{F}$ with $F \subseteq \phi_e$ for $e \in E(\mathbb{X}')$ (and g' remains to be assigned for faces with $e \in E(\mathbb{Y}')$). Consider such face $F \subseteq \phi_e = \phi_{(x,x')}$, then f' is already assigned on the boundary ∂F. If $\partial D(j) \subseteq \phi_e$ for $j \in W(\mathbb{Y}')$, then by construction, $\partial D(j)$ touches at least one component ($\partial D(e)$ or $\partial D(x')$) of $\partial\phi_e$. Therefore, each component of ∂F touches at least one component of $\partial\phi_e$. Moreover, because both $\partial D(e)$ and $\partial D(x')$ are connected, at most one component of ∂F touches $\partial D(e)$, and at most one component of ∂F touches $\partial D(x')$. Hence, ∂F has at most two components.

Consider the arcs of $\partial F \setminus \partial\phi_e$ and observe that, because one can draw at most one path between the boundaries of an annulus without disconnecting it, at most two arcs connect $\partial D(e)$ and $\partial D(x')$. Therefore, ∂F has either zero or two arcs on which f' is not constant. Similarly, g' is not constant for zero or two arcs of $\partial F \setminus \partial\phi_{E(\mathbb{Y}')}$, and $F \subseteq \phi_{E(\mathbb{Y}')}$ if such arcs exist.

Based on this, we distinguish the following four cases, illustrated in Figure 8.

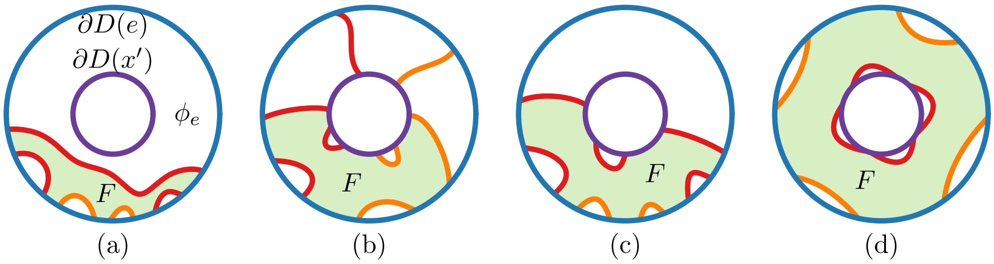

Figure 8: Possibilities for face $F \subseteq \phi_{e=(x,x')}$.

(a) ∂F touches at most one component of $\partial\phi_e$ (and hence ∂F has one component);

(b) ∂F has one component and touches both components of $\partial\phi_e$, and g is interpolated on both components of $\partial\phi_e$;

(c) ∂F has one component and touches both components of $\partial\phi_e$, and g is interpolated on ∂F on at most one component of $\partial\phi_e$;

(d) ∂F has two components (each touching a distinct component of $\partial\phi_e$).

In case (a), the value of f' is constant along ∂F, and we assign the same value for f' in the interior of F. We will see that the values of g' (which are assigned symmetrically) in the interior of F are values that are also assigned to g' on the boundary of F, so we have $\Delta(F) \leq \Delta(\partial F)$.

Consider the cases (b) and (c), where ∂F has two (interpolated) arcs, call them α_0 and α_1, both starting on $\partial D(e)$ and ending on $\partial D(x')$. We find two simple non-intersecting paths β_x and $\beta_{x'}$ in F, where β_x connects the start points of α_0 and α_1, and $\beta_{x'}$ connects the end points of α_0 and α_1. We assign $f(x)$ to f' on β_x, and $f(x')$ to f' on $\beta_{x'}$. The paths β_x and $\beta_{x'}$ divide F into three components: a component C_x with $f'(p) = f(x)$ for $p \in \partial C_x$, a component $C_{x'}$ with $f'(p) = f(x')$ for $p \in \partial C_{x'}$, and finally a component C_e bounded by α_0, α_1, β_x and $\beta_{x'}$.

We assign x for f' on the interior of C_x, x' for f' on the interior of $C_{x'}$, and interpolate C_e using the second argument of an arbitrary parameterization $\pi\colon [0,1] \times [0,1] \to C_e$, with $\pi(0,t) = \alpha_0(t)$, $\pi(1,t) = \alpha_1(t)$, $\pi(\cdot,0) = \beta_x$ and $\pi(\cdot,1) = \beta_{x'}$. See Figure 9.

If g' has a constant value for F, then clearly $\Delta(F) \leq \Delta(\partial F)$, so assume $F \subseteq \phi_{e'}$ for $e' = (y,y') \in E(\mathbb{Y}')$. In case (b), we have $\Delta(F) \leq \Delta(\partial F)$ because each of $(f(x),g(y))$, $(f(x'),g(y))$, $(f(x),g(y'))$, and $(f(x'),g(y'))$ appear as values for $(f'(p),g'(p))$ for points p on ∂F, so interpolation will not exceed the maximum difference. To ensure $\Delta(F) \leq \Delta(\partial F)$ in case (b), we must be careful in choosing C_e which interpolates f' and at the same time $C_{e'}$ which interpolates g' in F. Since in case (b), interpolation of f' occurs on a single component of $\partial\phi_{e'}$, and interpolation of g' occurs on a single component of g', we can choose β_0 and β_1 in such a way that C_e does not intersect $C_{e'}$, such that the interpolation of f' occurs in a region where g' is constant and vice-versa. Doing so, differences between f' and g' never exceed their difference at the boundary, so $\Delta(F) \leq \Delta(\partial F)$.

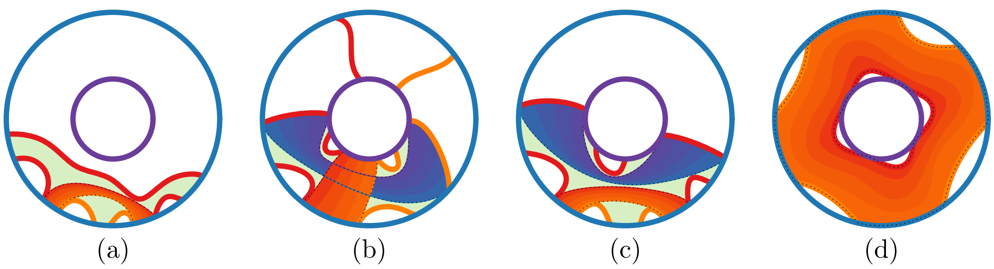

Figure 9: Disks C_e (and $C_{e'}$) interpolating f' (and g') in F for the four cases of Figure 8.

Lastly consider case (d), where F is an annulus, and on each boundary component, f' and g' are constant. By interpolating f' and g' between their value on each boundary component using the second argument of a parameterization $\pi\colon S^1 \times [0,1] \to F$, we ensure $\Delta(F) \leq \Delta(\partial F)$.

Theorems 3 and 4 follow.

Theorem 3 *Deciding whether the Fréchet distance between two $\mathbb{R}$-valued disks with constant values at their boundaries is at most ε is in NP.*

Proof. Observe that for $x \in V(\mathbb{X}')$, ϕ_x has the constant value $f(x)$ for f'. Moreover, for each $e = (x, x') \in E(\mathbb{Y}')$, ϕ_e interpolates f' between $f(x)$ and $f(x')$, so $d_F(f, f') = 0$. Similarly, $d_F(g', g) = 0$. Hence, if a topological expression on $W(\mathbb{X}') \cup W(\mathbb{Y}')$ respecting ε-matching τ is satisfiable, then $d_F(f, g) \leq \varepsilon$. Moreover, because such a topological expression exists for any ε-matching μ, deciding whether $d_F(f, g) < \varepsilon$ is in NP. □

Theorem 4 *Deciding whether the Fréchet distance between two $\mathbb{R}$-valued spheres is at most ε is in NP.*

Proof. Consider an ε-matching between the two spheres, and puncture one sphere at an arbitrary point, and puncture the other sphere at the matched point. The resulting surfaces are disks with constant values at their boundaries, and by Theorem 3 we can test in NP whether an ε-matching for the disks exists, in which case there is also an ε-matching between the spheres (because the boundaries of the disks must be aligned). □

3.4 NP-hardness

The contour tree distance is closely related to the Fréchet distance between surfaces because every $\mathbb{R}$-valued tree is the contour tree of some $\mathbb{R}$-valued surface. To construct such a surface from a contour tree, replace each edge of the tree by a cylinder and each degree d vertex by a sphere with d holes (one for each edge). The resulting surface is a topological sphere, representable as a simplicial 2-complex with linearly many triangles. Consider f and $g\colon S^2 \to \mathbb{R}$ to be $\mathbb{R}$-valued spheres with contour trees $\mathbb{X}$ and $\mathbb{Y}$.

Lemma 6 *Any many-to-one matching $\tau\colon \mathbb{X} \to \mathbb{Y}$ between contour trees of $\mathbb{R}$-valued spheres f and g can be lifted into an orientation preserving matching $\mu\colon S^2 \to S^2$, satisfying $\mathcal{R}_g \circ \mu = \tau \circ \mathcal{R}_f$.*

Proof. Assume τ is a many-to-one matching and denote by $C_f(x) = \mathcal{R}_f^{-1}(x)$ the contours in the preimage of $x \in \mathbb{X}$ under the quotient map $\mathcal{R}_f$ of the contour tree. For any degree d point $y \in \mathbb{Y}$, consider its preimage $X = \tau^{-1}(y)$. There are d components in $\mathbb{X} \setminus X$. We define μ to match $C_f(T)$ and $C_g(y)$ (both homeomorphic to d-holed spheres) by matching the boundary components as given by τ, so the interior of $C_f(T)$ is automatically matched to the interior of $C_g(y)$. Because τ is continuous, and $\tau^{-1}(y)$ and $\tau^{-1}(y')$ have an empty intersection for distinct y and $y' \in \mathbb{Y}$, this defines an orientation preserving homeomorphism μ between surfaces f and g. □

Corollary 2 *If τ is a many-to-one ε-matching between the contour trees $\mathbb{X}$ and $\mathbb{Y}$ of $\mathbb{R}$-valued spheres, then $d_F(f, g) \leq \varepsilon$.*

To prove NP-hardness for surfaces, we reuse the construction of the contour tree distance. Take $\mathbb{X}$ and $\mathbb{Y}$ to be the trees with functions f and g as constructed in Section 2.3, and construct $\mathbb{R}$-valued spheres with these trees as contour tree. To illustrate, f and g might be the contour trees of the surfaces depicted in Figure 10.

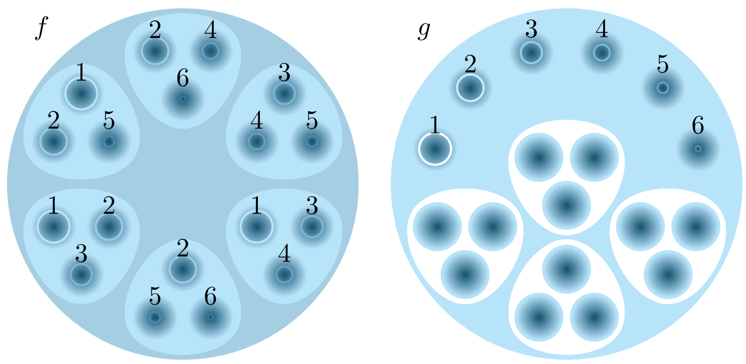

Figure 10: Two surfaces with color-coded function values. These surfaces are the result of the reduction from X3C with input $\mathcal{S} = \{\{2,4,6\},\{3,4,5\},\{1,3,4\},\{2,5,6\},\{1,2,3\},\{1,2,5\}\}$.

Theorem 5 *Computing a factor* $(2-\varepsilon)$*-approximation of the Fréchet distance between* $\mathbb{R}$*-valued spheres is NP-hard.*

Proof. By Lemma 5 and Corollary 1, $d_F(f,g) \geq d_C(f,g) \geq 2$ if $\mathcal{S}$ admits no solution to X3C. By Lemma 4 and Corollary 2, $d_F(f,g) \leq d_C(f,g) \leq 1$ if $\mathcal{S}$ admits a solution to X3C. So approximating the $d_F(f,g)$ for $\mathbb{R}$-valued spheres within factor 2 is NP-hard. □

Because the matching constructed in Lemma 4 always matched the two connection vertices, we can puncture the constructed surfaces at these vertices to obtain Corollary 3.

Corollary 3 *Computing a factor* $(2-\varepsilon)$*-approximation of the Fréchet distance between* $\mathbb{R}$*-valued disks is NP-hard.*

4 An unrepresentable matching

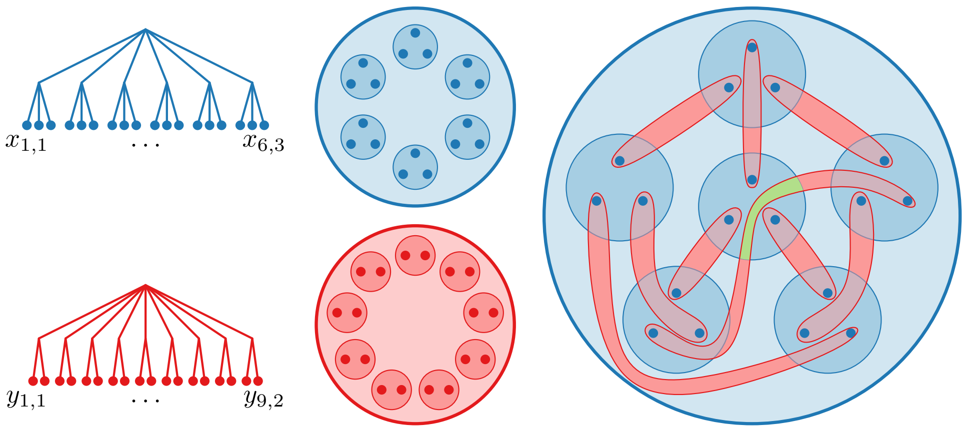

Figure 11: Left: trees $\mathbb{X}$ and $\mathbb{Y}$; middle: the corresponding surfaces; right: a matching in which a subtree of $\mathbb{Y}$ must intersect an additional subtree of $\mathbb{X}$ (highlighted in green).

Consider the two rooted trees $\mathbb{X}$ and $\mathbb{Y}$ in Figure 11. The leaves of $\mathbb{X}$ are labeled $x_{i,j}$ ($i \in \{1,\ldots,6\}$, $j \in \{1,2,3\}$) and the leaves of $\mathbb{Y}$ are labeled $y_{k,l}$ ($k \in \{1,\ldots,9\}$, $l \in \{1,2\}$). For both trees, leaves with the same i or k are grouped in subtrees. Based on the complete bipartite graph $K_{3,3}$ with vertices $v_1,\ldots,v_6$ and edges $e_1,\ldots,e_9$, we construct a matching τ between

those subtrees as follows. For an edge $e_k = (v_i, v_{i'})$ of $K_{3,3}$, match the path from $y_{k,1}$ to $y_{k,2}$ with the path between unused vertices $x_{i,j}$ and $x_{i',j'}$. Match the edge from the root to group i of $\mathbb{X}$ with the edges of $\mathbb{Y}$ from the root to the three groups that match with $x_{i,1}$, $x_{i,2}$ and $x_{i,3}$. Edges in the top layer of the tree are matched in a many-to-many fashion, whereas edges of the bottom layer are matched using linear interpolation. Then $\tau \in \mathcal{M}(\mathbb{X}, \mathbb{Y})$ does not match any path from $y_{k,1}$ to $y_{k,2}$ (of edge $e_k = (v_i, v_{i'})$) with any group of $\mathbb{X}$ not containing any $x_{i,j}$ or $x_{i',j'}$. However, because $K_{3,3}$ is not a planar graph, this matching cannot be realized on the sphere, as illustrated in Figure 11.
It therefore seems that the Fréchet distance between $\mathbb{R}$-valued spheres is more discriminative than the contour tree distance. However, this example merely shows that we cannot lift any ε-matching between contour trees to one between spheres; a different matching on spheres (such as in Figure 11) can still be an ε-matching. We have not found any instances (on spheres) for which the Fréchet distance is strictly greater than the contour tree distance.

5 Related measures

While the Fréchet distance is applicable to a broad class of functions, the functional distortion distance and interleaving distance specialize on Reeb graphs, and can be used to compare $\mathbb{R}$-valued functions. They are closely related to the Gromov-Hausdorff distance. The functional distortion distance $d_D(f, g)$ between Reeb graphs $f\colon \mathbb{X} \to \mathbb{R}$ and $g\colon \mathbb{Y} \to \mathbb{R}$ is defined as

$$\begin{aligned} d_D(f,g) &= \inf_{\Phi,\Psi} \max\{D(\Phi,\Psi), \|f - g\circ\Phi\|_\infty, \|g - f\circ\Psi\|_\infty\}, \text{ with} \\ D(\Phi,\Psi) &= \sup_{(x,y),(x',y')\in C(\Phi,\Psi)} \frac{1}{2}|\mathrm{height}_f(x,x') - \mathrm{height}_g(y,y')|, \text{ and} \\ C(\Phi,\Psi) &= \{(x,y)\in \mathbb{X}\times\mathbb{Y} \mid \Phi(x) = y \text{ or } x = \Psi(y)\}, \end{aligned}$$

where $\Phi\colon \mathbb{X} \to \mathbb{Y}$ and $\Psi\colon \mathbb{Y} \to \mathbb{X}$ are continuous maps, and

$$\mathrm{height}_f(x,x') = \inf_{\pi\colon x \rightsquigarrow_{\mathbb{X}} x'} \max_{x''\in\pi} f(x'') - \min_{x''\in\pi} f(x'')$$

is the length of the image (which is an interval) under f of the path in $\mathbb{X}$ from x to x' that minimizes this length. In the case of trees, this will be that of the unique simple path connecting x and x'.

Another measure is the weak Fréchet distance d_{wF}, which requires matchings to be surjective but—in contrast to the Fréchet distance—does not require the matching to be a homeomorphism. As such, it can compare trees that are not homeomorphic.

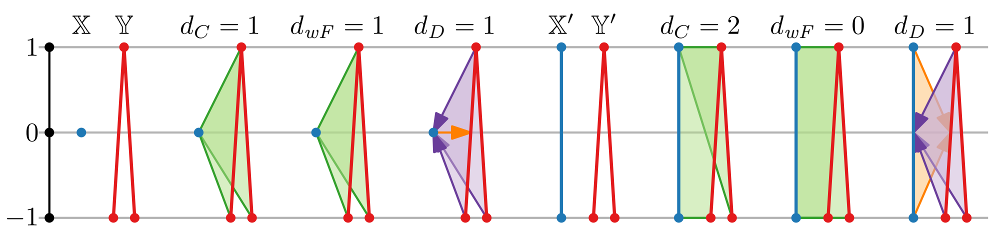

Figure 12: A comparison of the contour tree distance d_C, weak Fréchet distance d_{wF} and functional distortion distance d_D for small trees. Correspondences in green, Φ in orange, Ψ in purple.

We highlight how the functional distortion distance, the weak Fréchet distance and the contour tree distance differ using small instances in Figure 12. Perhaps surprising is that the functional distortion distance between $\mathbb{X}'$ and $\mathbb{Y}'$ is 1 instead of 2. This is because the term $D(\Phi, \Psi)$ is only half the difference in height between pairs of matched points, so by ensuring the difference in function value of matched points is at most half the difference in height between any two points, we achieve a distance of 1. It therefore seems that the contour tree distance is more discriminative than the functional distortion distance, and both are much more discriminative than the weak Fréchet distance.

6 Discussion

We have shown that computing the Fréchet distance between $\mathbb{R}$-valued surfaces is NP-complete if the surfaces are spheres or disks with constant boundary. The question whether the Fréchet distance is in NP for higher genus surfaces remains open. Using our techniques, the main bottleneck for this is that the string graph recognition problem is only known to lie in NEXP on higher genus surfaces [32], and it is therefore unknown whether Euler diagrams on arbitrary surfaces can be recognized in NP. So extensions of the techniques used in this paper might only show that the Fréchet distance is in NEXP for $\mathbb{R}$-valued surfaces of higher genus. Finally, we are interested in an efficient constant-factor approximation algorithm for the Fréchet distance or the contour tree distance, and whether these measures are equivalent on $\mathbb{R}$-valued spheres.

References

[1] Pankaj K. Agarwal, Rinat Ben Avraham, Haim Kaplan, and Micha Sharir. Computing the discrete Fréchet distance in subquadratic time. In *Proc. 24th Sympos. Discrete Algorithms (SODA)*, pages 156–167, 2013.

[2] Pankaj K. Agarwal, Kyle Fox, Abhinandan Nath, Anastasios Sidiropoulos, and Yusu Wang. Computing the Gromov-Hausdorff distance for metric trees. In *Proc. 26th Internat. Sympos. Algorithms Comput. (ISAAC)*, 2015.

[3] Alfred V. Aho, John E. Hopcroft, and Jeffrey D. Ullman. *The Design and Analysis of Computer Algorithms*. Addison-Wesley, 1974.

[4] Helmut Alt and Maike Buchin. Can we compute the similarity between surfaces? *Discrete Comput. Geom. (DCG)*, 43(1):78–99, 2010.

[5] Helmut Alt and Michael Godau. Computing the Fréchet distance between two polygonal curves. *Internat. J. Comput. Geom. Appl. (IJCGA)*, 5(01n02):75–91, 1995.

[6] Ulrich Bauer, Xiaoyin Ge, and Yusu Wang. Measuring distance between reeb graphs. In *Proc. 30th Sympos. Comput. Geom. (SoCG)*, page 464, 2014.

[7] Ulrich Bauer, Elizabeth Munch, and Yusu Wang. Strong equivalence of the interleaving and functional distortion metrics for Reeb graphs. In *Proc. 31st Sympos. Comput. Geom. (SoCG)*, pages 461–475, 2015.

[8] Kenes Beketayev, Damir Yeliussizov, Dmitriy Morozov, Gunther H. Weber, and Bernd Hamann. Measuring the distance between merge trees. In *Topological Methods in Data Analysis and Visualization III*, pages 151–165, 2014.

[9] Sotiris Brakatsoulas, Dieter Pfoser, Randall Salas, and Carola Wenk. On map-matching vehicle tracking data. In *Proc. 31st Int. Conf. Very Large Data Bases*, pages 853–864, 2005.

[10] Karl Bringmann. Why walking the dog takes time: Fréchet distance has no strongly subquadratic algorithms unless SETH fails. In *Proc. 55th IEEE Sympos. Found. Comput. Sci. (FOCS)*, pages 661–670, 2014.

[11] Karl Bringmann and Wolfgang Mulzer. Approximability of the discrete Fréchet distance. In *Proc. 31st Sympos. Comput. Geom. (SoCG)*, pages 739–753, 2015.

[12] Kevin Buchin, Maike Buchin, and Joachim Gudmundsson. Constrained free space diagrams: a tool for trajectory analysis. *Int. J. of GIS*, 24(7):1101–1125, 2010.

[13] Kevin Buchin, Maike Buchin, Wouter Meulemans, and Wolfgang Mulzer. Four soviets walk the dog-with an application to Alt's conjecture. In *Proc. 25th Sympos. Discrete Algorithms (SODA)*, pages 1399–1413, 2014.

[14] Kevin Buchin, Maike Buchin, and André Schulz. Fréchet distance of surfaces: Some simple hard cases. In *Proc. 18th European Sympos. Algorithms (ESA)*, pages 63–74, 2010.

[15] Kevin Buchin, Maike Buchin, Rolf van Leusden, Wouter Meulemans, and Wolfgang Mulzer. Computing the Fréchet distance with a retractable leash. In *Proc. 21st European Sympos. Algorithms (ESA)*, pages 241–252, 2013.

[16] Kevin Buchin, Maike Buchin, and Carola Wenk. Computing the Fréchet distance between simple polygons. *Comput. Geom. Theory Appl. (CGTA)*, 41(1–2):2–20, 2008.

[17] Kevin Buchin, Tim Ophelders, and Bettina Speckmann. Computing the similarity between moving curves. In *Proc. 23rd European Sympos. Algorithms (ESA)*, pages 928–940, 2015.

[18] Maike Buchin. *On the Computability of the Fréchet Distance Between Triangulated Surfaces*. PhD thesis, Free University Berlin, Institute of Computer Science, 2007.

[19] Atlas F. Cook IV, Anne Driemel, Sariel Har-Peled, Jessica Sherette, and Carola Wenk. Computing the Fréchet distance between folded polygons. In *Algorithms and Data Structures*, pages 267–278, 2011.

[20] Barbara Di Fabio and Claudia Landi. The edit distance for reeb graphs of surfaces. *Discrete Comput. Geom.*, 55(2):423–461, March 2016.

[21] Anne Driemel, Sariel Har-Peled, and Carola Wenk. Approximating the Fréchet distance for realistic curves in near linear time. *Discrete Comput. Geom. (DCG)*, 48(1):94–127, 2012.

[22] Max J. Egenhofer. Reasoning about binary topological relations. In *Sympos. on Spatial Databases*, pages 141–160, 1991.

[23] Maurice Fréchet. Sur quelques points du calcul fonctionnel. *Rendiconti Circ. Mat. Palermo*, 22:1–74, 1906.

[24] Maurice Fréchet. Sur la distance de deux surfaces. *Ann. Soc. Polon. Math.*, 3:4–19, 1924.

[25] Michael R. Garey and David S. Johnson. Computers and intractability: a guide to the theory of NP-completeness. *WH Freeman & Co., San Francisco*, 1979.

[26] Michael Godau. *On the complexity of measuring the similarity between geometric objects in higher dimensions*. PhD thesis, Freie Universität Berlin, Germany, 1998.

[27] Daniel Keysers and Walter Unger. Elastic image matching is NP-complete. *Pattern Recognition Letters*, 24(1):445–453, 2003.

[28] Dmitriy Morozov, Kenes Beketayev, and Gunther Weber. Interleaving distance between merge trees. *Discrete and Computational Geometry*, 49:22–45, 2013.

[29] Amir Nayyeri and Anastasios Sidiropoulos. Computing the Fréchet distance between polygons with holes. In *Proc. 42nd Internat. Colloq. Automata Lang. Program. (ICALP)*, pages 997–1009, 2015.

[30] Amir Nayyeri and Hanzhong Xu. On Computing the Fréchet Distance Between Surfaces. In *Proc. 32nd Sympos. Comput. Geom. (SoCG)*, volume 51, pages 55:1–55:15, 2016.

[31] Georges Reeb. Sur les points singuliers dune forme de Pfaff complètement intégrable ou dune fonction numérique. *C. R. de LAcadémie des Sciences*, 222:847–849, 1946.

[32] Marcus Schaefer, Eric Sedgwick, and Daniel Štefankovič. Recognizing string graphs in NP. *Journal of Computer and System Sciences*, 67(2):365 – 380, 2003.

[33] Marcus Schaefer and Daniel Štefankovič. Decidability of string graphs. In *Proc. 33rd Sympos. Theory Comput. (STOC)*, pages 241–246, 2001.

[34] Aristeidis Sotiras, Christos Davatzikos, and Nikos Paragios. Deformable medical image registration: A survey. *IEEE Transactions on Medical Imaging*, 32(7):1153–1190, 2013.

[35] Marc Van Kreveld, René van Oostrum, Chandrajit Bajaj, Valerio Pascucci, and Dan Schikore. Contour trees and small seed sets for isosurface traversal. In *Proc. 13th Sympos. Comput. Geom. (SoCG)*, pages 212–220, 1997.

[36] Daniel Štefankovič. *Algorithms for simple curves on surfaces, string graphs, and crossing numbers.* PhD thesis, University of Chicago, 2005.