

The Stability of the Apparent Contour of an Orientable 2-Manifold ^{*}

Herbert Edelsbrunner^{1,2,3}, Dmitriy Morozov⁴, and Amit Patel^{1,2}

¹ Dept. Comput. Sci., Duke Univ., Durham, North Carolina, USA.

² IST Austria (Institute of Science and Technology Austria).

³ Geomagic, Research Triangle Park, North Carolina, USA.

⁴ Dept. Comput. Sci. and Math, Stanford Univ., Stanford, California, USA.

Abstract. The (*apparent*) *contour* of a smooth mapping from a 2-manifold to the plane, $f : \mathbb{M} \rightarrow \mathbb{R}^2$, is the set of critical values, that is, the image of the points at which the gradients of the two component functions are linearly dependent. Assuming \mathbb{M} is compact and orientable and measuring difference with the erosion distance, we prove that the contour is stable.

1 Introduction

The most familiar setting of the problem studied in this paper is the view of a three-dimensional, solid body. We only see its surface and only one side at a time, but we get cues about its shape from the curve of points at which the surface normal is orthogonal to the viewing direction [17]. The view is the projection to a plane and its apparent contour is the image of the mentioned curve under the projection. Common roughly synonymous terms are *fold*, *silhouette*, *outline*, and *profile*. Only the first of these terms has a precise meaning introduced by Whitney [20]. Specifically, he defines *fold points* and *cusplike points* that admit parametrizations of the neighborhood such that the mapping can locally be written as $f(x_1, x_2) = (x_1^2, x_2)$ and $f(x_1, x_2) = (x_1(x_1^2 - x_2), x_2)$, and he showed that these are the only kinds of critical points that are stable under infinitesimal perturbations. We will refer to them as *double points* and *triple points* of the mapping. A related concept is the Jacobi curve as introduced in [9]. This is the set of critical points, and its image is the apparent contour. In computer graphics, the contour of the projection of a surface is often used for artistic enhancements of displays [6, 7]. In the typical case, the computational cost of the contour is significantly smaller than that of the entire surface [13, 16]. This motivates its use in efficient rendering; see [15] for a survey of algorithms generating contours. Additional applications are for shadow calculations, occlusion testing, and the simplification of surface models [8, 19].

The main result in this paper is a quantitative contribution to the structural stability of the apparent contour. This study began with Whitney's seminal paper [20] which originated the related fields of catastrophe theory [2] and singularity theory [14]. Looking at smooth mappings of manifolds, these fields focus on the structure of singularities

^{*} This research is partially supported by the Defense Advanced Research Projects Agency (DARPA) under grants HR0011-05-1-0007 and HR0011-05-1-0057.

and their stability under infinitesimal perturbations. In contrast to these studies, we allow for more severe perturbations and we quantify the changes in the contour. In a nutshell, we prove that a small perturbation of the mapping of an orientable 2-manifold to the plane changes the apparent contour only slightly. This seems plausible but is false for naive measurements of the distance between two contours. Indeed, even small perturbations can introduce arbitrarily many creases, each delimited by arbitrarily long contour lines. The crucial insight is that creases are thin, that is, they are delimited by pairs of contour lines that run roughly parallel to each other at close distance. We thus have two cases: creases that are thin in a technical sense that we will make precise shortly, and contour lines that are close to contour lines of the unperturbed mapping. This distinction is crucial in any effort to simplify the contour of a mapping in a way that retains its essential character.

Outline. Section 2 explains the setting for our study. Section 3 introduces important concepts. Section 4 presents our main result, a global statement of stability of the apparent contour. Section 5 contains the proof. Section 6 concludes the paper.

2 The Setting

In this section, we describe the setting, namely generic, smooth mappings from an orientable 2-manifold to the plane.

The apparent contour. Instead of projections of surfaces, we consider the more general setting of mappings $f : \mathbb{M} \rightarrow \mathbb{R}^2$, in which \mathbb{M} is a compact, orientable 2-manifold without boundary. It may or may not be embedded in \mathbb{R}^3 . We assume the mapping is smooth and satisfies a small number of requirements we need in the proof of our result. This includes that for most points $x \in \mathbb{M}$, the derivative of f at x , which we denote as $Df(x) : \mathbb{R}^2 \rightarrow \mathbb{R}^2$, is surjective. We call these the *regular points* of f . All other points of \mathbb{M} are *critical points* of f and their images are *critical values*. A point in \mathbb{R}^2

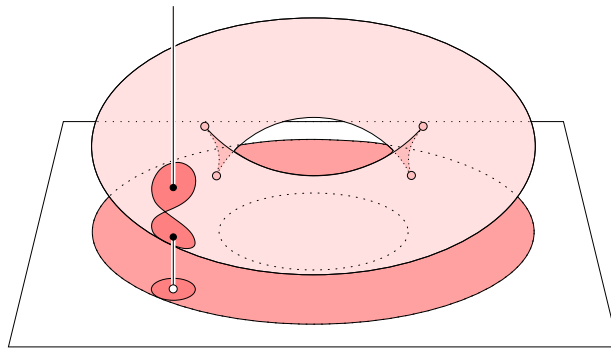


Fig. 1. The projection of the torus to the plane. The distance function defined by the marked value in the plane is illustrated by showing one of its sublevel sets.

is a *regular value* if it is not a critical value or, equivalently, if all its preimages are regular points. We call the set of critical values the (*apparent*) *contour* of the mapping, denoting it by $\text{Contour}(f)$. The adjective serves as a reminder that we are not talking about a structural property of the 2-manifold but rather of its mapping to the plane. We refer to Figure 1 for an example. The vertical projection of the torus to the plane below it has a contour that consists of two concentric circles. In contrast, the projection to the drawing plane is more complicated, with the inner circle twisting up twice, forming a triangle on the left and another one on the right. Two vertices of each triangle are *cusps*, that is, values at which the contour comes to a sudden halt and reverses its direction. The third vertex is a *crossing*, that is, a value at which the contour intersects itself.

Distance functions. Fixing a value $a \in \mathbb{R}^2$, we let $f_a : \mathbb{M} \rightarrow \mathbb{R}$ be the *distance function* that maps every point x to the Euclidean distance of its image from a , that is, $f_a(x) = \|f(x) - a\|_2$. In Figure 1, one such distance function is illustrated by showing the value, the two points in its preimage, a disk around the value, and the preimage of the disk. The *sublevel set* of f_a for threshold $r \geq 0$ is the set of points with function value at most r , denoted as $\mathbb{M}_r(a) = f_a^{-1}[0, r]$. Writing $B_r(a)$ for the closed disk with center $a \in \mathbb{R}^2$ and radius r , the sublevel set of f_a is the preimage of this disk, $\mathbb{M}_r(a) = f^{-1}(B_r(a))$. Following Morse theoretical ideas, we increase r and notice that the sublevel set changes its topology only at critical values. To explore this, we apply the chain rule to get the derivative of f_a at a point x as the composition of the derivative of f at x and the scalar product with $u = f(x) - a$. Writing $Df(x) : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ for the former and $\sigma_u : \mathbb{R}^2 \rightarrow \mathbb{R}$ for the latter, we have $Df_a(x) = \sigma_u \circ Df(x) : \mathbb{R}^2 \rightarrow \mathbb{R}$. The point x is regular for f_a if $Df_a(x)$ is surjective and it is critical if $Df_a(x)$ is the zero function. If x is regular for f then $Df(x)$ is surjective, and unless $f(x) = a$, this implies that $Df_a(x)$ is surjective and hence x is regular for f_a . On the other hand, if x is critical for f then it may or may not be critical for f_a but it will be critical for the distance function defined by another value.

CRITICAL POINT LEMMA. A point $x \in \mathbb{M}$ is critical for f iff there exists a value $a \neq f(x)$ in \mathbb{R}^2 such that x is critical for f_a .

Genericity. We aim at limiting the class of mappings to those with manageable properties. For reasons of exposition, we do not strive to make the class as large as possible but rather large enough to be interesting. In particular, we sacrifice some generality to avoid the need to explain homology groups before giving the proofs.

DEFINITION. Let \mathbb{M} be a compact, orientable 2-manifold without boundary. We call a smooth mapping $f : \mathbb{M} \rightarrow \mathbb{R}^2$ *generic* if

- (I) the distance function, f_a , is tame for every value $a \in \mathbb{R}^2$;
- (II) there are no critical points of f beyond double and triple points;
- (III) the apparent contour of f has finitely many cusps and crossings.

Condition (I) means that f_a has only a finite number of critical values and every sublevel set consists only of finitely many components with finitely many holes. Condition (II) prohibits critical points other than the two simple types, fold points and cusp points.

Not allowing other, more complicated critical points is convenient and not a serious restriction since the corresponding mappings are dense in the larger class of smooth mappings [20]. Condition (III) implies the existence of a finite, smooth stratification of the plane compatible with the contour. Specifically, with $\mathbb{X}^1 = \text{Contour}(f)$ and \mathbb{X}^0 the set of cusps and crossings, we have a decomposition $\emptyset = \mathbb{X}^{-1} \subseteq \mathbb{X}^0 \subseteq \mathbb{X}^1 \subseteq \mathbb{X}^2 = \mathbb{R}^2$ such that each *stratum*, $S^i = \mathbb{X}^i - \mathbb{X}^{i-1}$, is either empty or an i -dimensional manifold. Calling the components of the strata the *pieces* of the stratification, we have only a finite number and each piece is smoothly embedded in \mathbb{R}^2 . We call the pieces in S^2 the *chambers* of the stratification.

3 The Concepts

In this section, we introduce the main concepts needed to give a precise statement of our result.

Degree. An important step in our construction is the assessment of the shape of a component in the sublevel set of a distance function. For this purpose, we use the standard concept of degree, which we first define for the entire 2-manifold. To begin, we orient all sufficiently small, simple, closed curves in \mathbb{M} in a consistent manner. This is possible because \mathbb{M} is orientable. Let $a \in \mathbb{R}^2$ be a regular value and $f(x) = a$. A sufficiently small, simple, closed curve going around x in \mathbb{M} maps injectively to a closed curve going around a in \mathbb{R}^2 . We count +1 if this curve goes around a in a counterclockwise order and -1 if it goes around a in a clockwise order. Finally, we sum these numbers over all preimages of a and denote the result as $\deg(f, a)$. Extending this definition to critical values, we count 0 for each double point and +1 or -1 for each triple point. Moving a continuously from one value to another does not change this number. This is clear if we stay inside the same chamber but also when we cross the contour, picking up or losing two points whose contributions cancel each other. This implies $\deg(f, a) = \deg(f, b)$ for all $a, b \in \mathbb{R}^2$, and it makes sense to call this value the *degree* of f , denoted as $\deg(f)$. Since \mathbb{M} is compact, its image under f does not exhaust \mathbb{R}^2 . The degree of f at a value outside the image vanishes, which implies $\deg(f) = 0$.

Next, consider the closed disk $B_r(a)$ with center $a \in \mathbb{R}^2$ and radius $r \geq 0$. As mentioned earlier, the preimage of this disk is the sublevel set of f_a for threshold r . Let C be a component of this sublevel set and $f|_C : C \rightarrow \mathbb{R}^2$ the restriction of f to C . For each value b in $B_r(a)$, we get $\deg(f|_C, b)$ by summing the contributions over all points in the preimage, $f|_C^{-1}(b) = f^{-1}(b) \cap C$. As before, this number is the same at all values in the disk so we can call it the *degree* of $f|_C$. However, the number may change when we leave the disk so the degree is not necessarily zero.

Level sets and well function. We study the contour in terms of the family of preimages of all values. For a given value $a \in \mathbb{R}^2$, we call the preimage the *level set* of f at a . Equivalently, this is the zero set of the corresponding distance function, $f^{-1}(a) = f_a^{-1}(0)$. By Condition (I), this is a finite set of points in \mathbb{M} . The central concept in our approach is the health of these points. Considering the sublevel set, $\mathbb{M}_r(a) = f_a^{-1}[0, r]$, we wish to distinguish between components that necessarily map

to the entire disk and components that can be pushed off the disk with moderate effort. We call a component C of $\mathbb{M}_r(a)$ *well* if the degree of $f|_C$ is non-zero and *ill* if the degree is zero. For $r = 0$, each point in the level set forms its own component, which is well if the point is regular or a cusp and ill if it is a double point. As we increase r , we get a nested sequence of sublevel sets, $\mathbb{M}_r(a) \subseteq \mathbb{M}_s(a)$ for all $0 \leq r \leq s < \infty$. If the interval $[r, s]$ does not contain any critical value of f_a then the components of $\mathbb{M}_r(a)$ grow continuously into those of $\mathbb{M}_s(a)$ without changing their degree and status. At a simple critical value of f_a , we either encounter a new component or we merge two old components into one. The newly formed component has vanishing degree and is therefore ill from the start. When we merge two components, we add their degrees. The status of the new component thus depends only on the status of the old components. Specifically, merging a well and an ill component gives a well component, while merging two well or two ill components gives an ill component. There are also non-simple critical values, where we encounter two or more critical points at the same time or we encounter a cusp approaching it from its normal direction. In such a case, the change in the level set can be understood as the composition of a few simple changes as described.

The history of a component in the nested sequence of sublevel sets is therefore straightforward. If the component begins as a regular point or a triple point then it starts out well and falls ill later, at some critical value of f_a . We call this a *terminal critical value* to distinguish it from others at which no component falls ill. If the component begins as a double point then it is ill from the start. Once a component is ill, it does not get well any more (except it can become part of another, well component). It thus makes sense to introduce a function $\varphi : \mathbb{M} \rightarrow \mathbb{R}$ defined by mapping x to the terminal critical value of f_a , with $a = f(x)$, at which the component that contains x falls ill. We call φ the *well function* of f and $\varphi(x)$ the *well threshold* of x . We have $\varphi(x) = 0$ iff x is a double point and $\varphi(x) > 0$ if x is a regular point or a triple point.

Well diagrams. Fixing a value $a \in \mathbb{R}^2$, we get a well threshold for each point $x \in f_a^{-1}(0)$. We collect these thresholds to form a multiset of real numbers, called the *well diagram* of f_a and denoted as $\text{Dgm}(f_a)$. For most regular values a of f , this diagram consists of an even number of positive thresholds that come in equal pairs. For each pair, the two contributing points lie on patches facing opposite directions. A threshold in $\text{Dgm}(f_a)$ is *simple* if it is positive and occurs only twice. A non-trivial property of the well diagram is its stability; see the Stability Theorem for Well Diagrams in Section 5 but also [11]. To make this precise, let $a, b \in \mathbb{R}^2$ and let $0 \leq u_1 \leq u_2 \leq \dots \leq u_l$ and $0 \leq v_1 \leq v_2 \leq \dots \leq v_l$ be the thresholds in the well diagrams of f_a and f_b , possibly after adding zeros to the shorter sequence so we get the same length for both. The mentioned theorem states that $|u_i - v_i|$ is bounded from above by the L_∞ -difference between the two functions. Using the triangle inequality, we get $\|f_a - f_b\|_\infty \leq \|a - b\|_2$ and therefore

$$\max_{1 \leq i \leq l} |u_i - v_i| \leq \|a - b\|_2. \quad (1)$$

In words, corresponding well thresholds change at most by the Euclidean distance between the values in the plane.

Surgery. The stability of the well diagram expressed in (1) provides some hope that the well function defined earlier is continuous. This is indeed the case at points with simple well thresholds. Similarly, φ is continuous at double points, where it is zero. However, continuity is not guaranteed at points with positive but non-simple well thresholds. Let $a \in \mathbb{R}^2$ be a value and $u > 0$ a threshold that occurs at least four times in $\text{Dgm}(f_a)$. Equivalently, a is equidistant to at least two different values of the contour. Take for

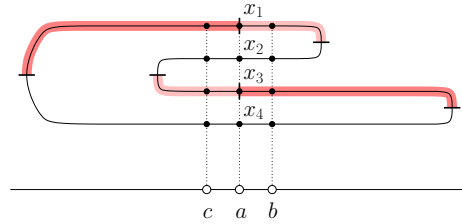


Fig. 2. Schematic cross-section of the twisted triangle in Figure 3. To make the well function continuous, we cut at x_1 and x_3 and reglue the sides as indicated by the light and dark shading.

example a value a somewhere in the interior of the bold, black segment connecting the upper left cusp with the center of the twisted triangle in Figure 3. It has four points in its preimage, x_1, x_2, x_3, x_4 , indexed from front to back in the picture. Figure 2 shows a section of the configuration, crossing the black, bold segment at a . These four points have four identical well thresholds. For a value b to the right of a , the top (front) two points in the preimage have a small well threshold, while for a value c to the left of a , the middle two points have a small well threshold. The other preimages of b and c have a large well threshold. As we move from b to c , we observe a jump of φ at x_1 and x_3 . The same jump occurs along the entire length of the black, bold segment. We remedy the discontinuity by cutting along the segment and regluing the sides as necessary to get continuity. In particular, the surfaces to the left of x_1 and to the right of x_3 are glued and so are the surfaces to the right of x_1 and to the left of x_3 .

Branch points. Even more interesting is what happens at the center, a , of the twisted triangle, the lower right endpoint of the bold segment in Figure 3. It has three closest values on the contour. We study the structure by going around a in a counterclockwise circle and drawing the well diagrams as we go. Each value b on this circle has four preimages, y_1, y_2, y_3, y_4 , indexed from front to back, as before. Growing the disk centered at b , we get a tree that describes how the components of the preimage merge until only one component remains. All four components start out well and fall ill in pairs during the process. This is illustrated in Figure 3, where well components are represented by bold branches in the trees and their falling ill is marked by shaded dots. As discussed earlier, there is a switch between y_1 and y_3 when we cross the bold segment. Symmetrically, there is a switch between y_2 and y_4 when we cross the segment connecting the lower cusp with the center of the triangle. The switches imply that we have

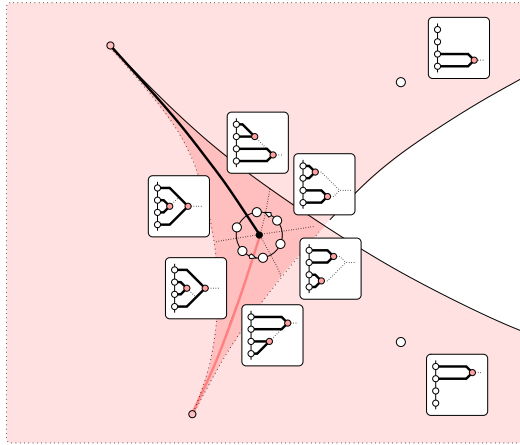


Fig. 3. Enlarged view of the twisted triangle to the left of the hole in the torus in Figure 1. The trees sketch the health histories of the points in the level sets at the marked values.

to go around the circle twice to return to the original configuration. In other words, the surgery along the bold segment creates a *branch point* at the center of the triangle, that is, a point with a disk neighborhood that covers the neighborhood of $a \in \mathbb{R}^2$ twice. Using complex numbers to parametrize the neighborhood of a , the map to the neighborhood of the branch point can locally be written as $z \mapsto z^2/\|z\|$.

We note that the situation leading to the creation of the branch point reminds us of the concept of a ring species in biology; see e.g. [12]. Locally, at a value b , we seem to have two distinct species, y_1 and y_3 , which we discover to be the same if we take a more global view of the situation.

Summary. Using the distance function defined for a value $a \in \mathbb{R}^2$, we have defined well thresholds for the points $x \in f^{-1}(a)$, and by exhausting all values in the plane, we have constructed a well function, $\varphi : \mathbb{M} \rightarrow \mathbb{R}$. Similar to the elevation function defined in [1], the well function is continuous almost everywhere but not necessarily everywhere. The stability of the well diagram implies that we can do surgery to change \mathbb{M} to a 2-manifold with boundary, Φ , on which the well function is continuous. Specifically, we cut \mathbb{M} along the curve of critical points of f . Doing so, we double every point to form the boundary of the 2-manifold with boundary. In addition, we cut and reglue along select curves originating at triple points. When we cut, we double the points and when we glue, we identify points in pairs. The two operations change the topology but cancel each other's effect on the multiplicity of points in the interior of the cut lines. Each such line starts at the third copy of a triple point (the first two copies are part of the boundary) and either ends at the third copy of another triple point or at a branch point. For reasons that will become clear later, we keep each branch point as two points with indistinguishable neighborhoods. The result is a non-Hausdorff 2-manifold with boundary, Φ , and a continuous well function, $\varphi : \Phi \rightarrow \mathbb{R}$. It vanishes along the boundary and is positive everywhere else.

4 The Result

In this section, we compare two mappings of the same 2-manifold and relate the difference between the contours to the difference between the mappings. We have two statements of stability. The first is straightforward and leads up to the second statement, our main result.

Silhouette stability. The *silhouette* of a mapping $f : \mathbb{M} \rightarrow \mathbb{R}^2$ is the boundary of the image, $\text{Sil}(f) = \text{bd im } f$. Thinking of the image as the foreground and its complement as the background, the silhouette is the subset of the contour that separates foreground from background. To compare the silhouette of f with that of another mapping $g : \mathbb{M} \rightarrow \mathbb{R}^2$, we define the *dilation* of a set $A \subseteq \mathbb{R}^2$ by a radius $\varepsilon \geq 0$ as the set of points in \mathbb{R}^2 at distance at most ε from some point in A . We denote this set by $A^{+\varepsilon}$. The *Hausdorff* or *dilation distance* between two sets $A, B \subseteq \mathbb{R}^2$ is the infimum of the radii for which each dilated set contains the other, un-dilated set,

$$D(A, B) = \inf\{\varepsilon \mid A \subseteq B^{+\varepsilon} \text{ and } B \subseteq A^{+\varepsilon}\}.$$

Setting $\varepsilon = \max_{x \in \mathbb{M}} \|f(x) - g(x)\|_2$, we can be sure that every value in the image of f has a value in the image of g at distance at most ε . Together with the symmetric relation, this implies our first result.

SILHOUETTE STABILITY LEMMA. The Hausdorff distance between the images of f and g is $D(\text{im } f, \text{im } g) \leq \max_{x \in \mathbb{M}} \|f(x) - g(x)\|_2$.

This result is nothing short of trivial and allows for easy generalizations to higher dimensions, spaces that are not manifolds, and mappings that are neither generic nor smooth. Note that the small Hausdorff distance between the images does not imply that the two silhouettes are everywhere close. Indeed, it allows for small holes arbitrarily far from the other silhouette.

Erosion distance. When we consider the entire contour then small holes cannot disappear without a trace. To the contrary, little islands may appear or disappear anywhere inside the foreground. This motivates us to define the *erosion* of a set $A \subseteq \mathbb{R}^2$ by a radius $\varepsilon \geq 0$ is obtained by removing all points at distance at most ε from the complement, that is, $A^{-\varepsilon} = \mathbb{R}^2 - (\mathbb{R}^2 - A)^{+\varepsilon}$. The *complementary Hausdorff* or *erosion distance* between two sets $A, B \subseteq \mathbb{R}^2$ is the infimum of the radii for which each eroded set is contained in the other, un-eroded set,

$$E(A, B) = \inf\{\varepsilon \mid A^{-\varepsilon} \subseteq B \text{ and } B^{-\varepsilon} \subseteq A\}.$$

To extend the idea of erosion to the manifold, we note a relation between the well function and the Euclidean distance in the image stated as the Well Function Lemma in Section 5. Specifically, $\varphi(x)$ is the distance between $a = f(x)$ and a locally closest value of $\text{Contour}(f)$. In other words, $\varphi(x)$ measures how far x is from the relevant portion of the boundary of Φ , and this measure is taken in the image rather than on the manifold. Eroding in the plane thus generalizes to taking a superlevel set of the well

function, that is, $\Phi^{-\varepsilon} = \varphi^{-1}[\varepsilon, \infty)$. Letting $g : \mathbb{M} \rightarrow \mathbb{R}^2$ be another generic, smooth mapping and $\gamma : \Gamma \rightarrow \mathbb{R}$ its well function after surgery, we define $\Gamma^{-\varepsilon} = \gamma^{-1}[\varepsilon, \infty)$. The *erosion distance* between Φ and Γ is then the infimum of the radii $\varepsilon \geq 0$ for which there are injections $\iota_f : \Phi^{-\varepsilon} \rightarrow \Gamma$ and $\iota_g : \Gamma^{-\varepsilon} \rightarrow \Phi$ such that $f(x) = g \circ \iota_f(x)$ and $g(y) = f \circ \iota_g(y)$ for all points x and y .

Contour stability. We are now ready to state the main result of this paper. It compares the image of the eroded 2-manifold, $\Phi^{-\varepsilon}$, using f , with the image of the un-eroded 2-manifold, Γ , using g , where $\varepsilon = \max_{x \in \mathbb{M}} \|f(x) - g(x)\|_2$, as before. Specifically, it says the second mapping covers every value in \mathbb{R}^2 at least as often as the first mapping. The same is true if we exchange f and g .

CONTOUR STABILITY THEOREM. Let \mathbb{M} be a compact, orientable 2-manifold without boundary and $f, g : \mathbb{M} \rightarrow \mathbb{R}^2$ two generic, smooth mappings. Then the erosion distance is $E(\Phi, \Gamma) \leq \max_{x \in \mathbb{M}} \|f(x) - g(x)\|_2$.

We illustrate the result in Figure 4, which shows the familiar projection of the torus superimposed on a perturbation of that projection. The perturbed mapping has two extra cusps connected to each other by two contour lines bounding a narrow lip-shaped chamber. Cutting along the corresponding curves of critical points, we get a hole in the surface, which we mend by cutting and regluing along two preimages of the medial line between the two contour lines of the lips; see Figure 4. We get no additional branch point but instead two new components, each covering the lips once.

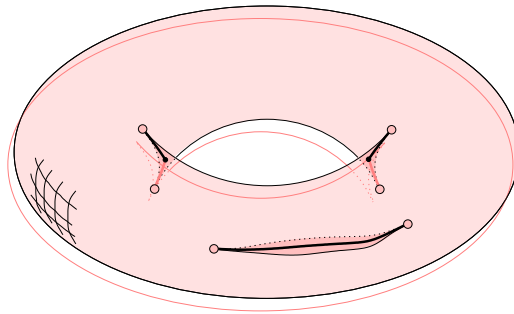


Fig. 4. Superposition of the faint contour of the original mappings of the torus and the clear contour of the perturbed mapping.

At this juncture, we wish to draw attention to the fact we use injections in the definition of the erosion distance. Write $E_{\text{strong}}(\Phi, \Gamma)$ for the strong version in which we require ι_f and ι_g be embeddings. Clearly, $E(\Phi, \Gamma) \leq E_{\text{strong}}(\Phi, \Gamma)$ so that substituting the strong for the original version of erosion distance would give a stronger theorem. Our proof does not support this strengthening. Although we currently do not have an example that shows such a strengthening is impossible, we believe such examples exist.

5 The Proof

In this section, we present the proof of our main result, delegating the bulk of the underlying algebraic construction to [11].

From components to homology groups. In lieu of the components in the sublevel set, $\mathbb{M}_r(a)$, we consider the 0-dimensional homology group of that set, which we denote as $F_r(a) = H_0(\mathbb{M}_r(a))$. With this formalization, we gain access to the concept of persistence, as introduced in [10]. Particularly important is the stability of the persistence diagram, which was established for tame functions in [5]. To explain this result, we consider again the nested sequence of sublevel sets, $\mathbb{M}_r(a) \subseteq \mathbb{M}_s(a)$ for $0 \leq r \leq s < \infty$. The inclusion between two sublevel sets induces a homomorphism between the corresponding homology groups, giving rise to $0 \rightarrow \dots \rightarrow F_r(a) \rightarrow F_s(a) \rightarrow \dots$, which we call a *filtration*. Within it, a component is *born* at $F_r(a)$ if the minimum function value of its points is r , and it *dies entering* $F_s(a)$ if it merges at s with another component born before itself. The component is thus characterized by two numbers, r and s , which we interpret as coordinates of a point in the plane. We set $s = \infty$ if the component never dies, so we need the extended plane, $\mathbb{R}^2 = [-\infty, \infty]^2$, to draw the points. Representing each component that ever appears in the filtration, we get a multiset in \mathbb{R}^2 , which we call the *persistence diagram* of f_a , denoted as $\text{Dgm}(f_a)$. For a technical reason that will be clear shortly, we add infinitely many copies of every point on the diagonal to the diagram. Letting $g : \mathbb{M} \rightarrow \mathbb{R}^2$ be a second mapping, we get a second distance function and a second persistence diagram, $\text{Dgm}(g_a)$. Using the triangle inequality, it is easy to show that the difference between the distance functions is $\|f_a - g_a\|_\infty \leq \varepsilon$, where $\varepsilon = \max_{x \in \mathbb{M}} \|f(x) - g(x)\|_2$. The mentioned stability result states that the bottleneck distance between the persistence diagrams is bounded by the difference between the functions and therefore by ε , that is,

$$W_\infty(\text{Dgm}(f_a), \text{Dgm}(g_a)) \leq \varepsilon, \quad (2)$$

see [5]. This means there is a perfect matching between the points in the two diagrams such that the L_∞ -distance between matched points is at most ε . This result suffices to derive a local statement of contour stability but not the stronger, global statement given in Section 4.

Equivalence of definitions. To go the extra mile, we need to understand the subgroups of the homology groups generated by the well components of the sublevel sets. We refer to these as the *well groups*, $U_r(a) \subseteq F_r(a)$. These groups have been studied in [11], where a different, more general definition is used. We reproduce this definition. Letting $f, h : \mathbb{M} \rightarrow \mathbb{R}^2$ be two mappings, we call h a ρ -*perturbation* of f if $\max_{x \in \mathbb{M}} \|h(x) - f(x)\|_2 \leq \rho$. Note that the level set of h at a is contained in the sublevel set of f_a for radius ρ , that is, $h^{-1}(a) \subseteq \mathbb{M}_\rho(a)$. Hence, there is a homomorphism $j_h : H_0(h^{-1}(a)) \rightarrow F_\rho(a)$. The image of j_h is a subgroup of $F_\rho(a)$ and so is the common intersection of like images, $\bigcap_h \text{im } j_h \subseteq F_\rho(a)$, where h ranges over all ρ -perturbations of f . Finally, we set $\rho = r + \delta$ for a sufficiently small $\delta > 0$, and we define $W_r(a)$ as the largest subgroup of $F_r(a)$ so its image in $F_\rho(a)$ is contained in this common intersection. The group $W_r(a)$ is what [11] calls the well group of $\mathbb{M}_r(a)$. Our

aim here is to prove that for the setting in this paper, the two definitions give the same groups.

WELL GROUP LEMMA. We have $U_r(a) = W_r(a)$ for every $a \in \mathbb{R}^2$ and every $r \geq 0$.

Proof. Fixing $a \in \mathbb{R}^2$, we consider a point $x \in f^{-1}(a)$, and for every $r \geq 0$, we let C_r be the component of $\mathbb{M}_r(a)$ that contains x .

CASE 1: C_r is well. We show that there exists $\delta > 0$ such that $C_{r+\delta} \cap h^{-1}(a) \neq \emptyset$ for every ρ -perturbation h of f , where $\rho < r + \delta$. Specifically, we choose $\delta < \varphi(x) - r$ and note that $C_{r+\delta}$ is well. Consider the homotopy defined by $g_t(x) = (1 - t)f(x) + th(x)$. Since the boundary of $C_{r+\delta}$ is too far from the center for its image to reach a , the degree of g_t restricted to $C_{r+\delta}$ at a remains unchanged. This degree is non-zero for $f = g_0$ and therefore also non-zero for $h = g_1$. This implies that $h^{-1}(a)$ has a non-empty intersection with $C_{r+\delta}$, as required.

CASE 2: C_r is ill. We show that for every $\delta > 0$ there exists a $\rho < r + \delta$ and a ρ -perturbation h of f such that $C_r \cap h^{-1}(a) = \emptyset$. We use induction, following the change in the sublevel set as we increase the radius. The first time we have to prove something is when $r = \varphi(x)$. At this radius, two well components merge to form C_r , which is now ill. Let y be a double point at which the two components touch; see Figure 5. The

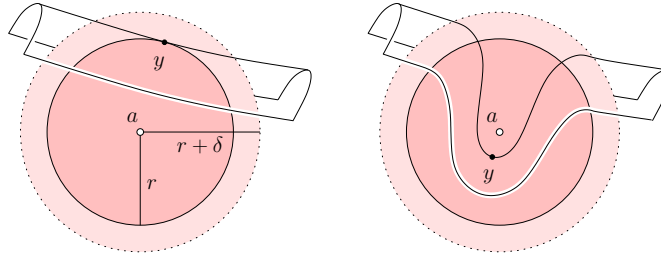


Fig. 5. Left: two well components meeting at y . Right: the locally perturbed mapping in which the two merged components avoid a .

perturbation needs to move y beyond a , which it can do without changing f outside $C_{r+\delta}$. If there are two or more such double points, we move all of them beyond a the same way. We choose $\delta < s - \varphi(x)$, where s is the next, larger critical value of f_a , and call the resulting perturbation $h_0 : \mathbb{M} \rightarrow \mathbb{R}^2$. It is good for all radii $\varphi(x) \leq r < s$. Now suppose $r = s$ and the growing component merges with another, ill component, forming C_s . Let $h_1 : \mathbb{M} \rightarrow \mathbb{R}^2$ be the perturbation we constructed for this other component when it fell ill at $\varphi(x') < s$. Choose t such that $\max\{\varphi(x), \varphi(x')\} < t < s$. The two perturbations differ from each other in two disjoint components of \mathbb{M}_t . We can therefore combine them to get a new perturbation $h_{01} : \mathbb{M} \rightarrow \mathbb{R}^2$ that agrees with f outside these components, with h_0 inside one, and with h_1 inside the other component. The level set of h_{01} at a has empty intersection with C_r , as required. The claimed relationship follows by induction. \square

Stability of diagram. While being more complicated algebraically, the persistence diagram of the well groups is simpler geometrically. Specifically, it is only 1-dimensional, namely precisely the well diagram introduced in Section 3. The complete proof of the stability of the well diagram is beyond the scope of this paper. The main idea is the realization that the well groups for a given value form a zigzag module as defined in [4]. We sketch the construction of this module for the distance function $f_a : \mathbb{M} \rightarrow \mathbb{R}$. By definition of a generic, smooth mapping, f_a has only finitely many critical values and therefore only finitely many different homology groups. We index them consecutively as F_i . Let $U_i \subseteq F_i$ be the corresponding well groups. A class may fall ill entering U_{i+1} because it dies entering F_{i+1} or because its image in F_{i+1} does not belong to U_{i+1} . To express the two cases algebraically, we let Q_i be the quotient formed by identifying all classes in U_i that differ only by a class that maps to zero in F_{i+1} . Inserting the quotient between the two well groups and connecting it with the obvious forward and backward maps, we get the zigzag module, $\dots \leftarrow U_i \rightarrow Q_i \leftarrow U_{i+1} \rightarrow \dots$. It is characterized by its persistence diagram, like a filtration [4]. By the Well Group Lemma, this diagram is precisely the well diagram described in Section 3. Stability does not follow from general principles known yet but has been established in [11]. We skip the argument and state the result.

STABILITY THEOREM FOR WELL DIAGRAMS. Let $f, g : \mathbb{M} \rightarrow \mathbb{R}^2$ be two generic, smooth mappings. Then the bottleneck distance between the well diagrams of the distance functions at any value $a \in \mathbb{R}^2$ is $W_\infty(\text{Dgm}(f_a), \text{Dgm}(g_a)) \leq \|f_a - g_a\|_\infty$.

As mentioned earlier, the difference between the distance functions is bounded from above by $\varepsilon = \max_{x \in \mathbb{M}} \|f(x) - g(x)\|_2$.

Eroding the manifold. The stability of the well diagram justifies the surgery which turns \mathbb{M} into a non-Hausdorff 2-manifold with boundary, Φ , such that the well function, $\varphi : \Phi \rightarrow \mathbb{R}$, is continuous. We recall that for each point $x \in \Phi$, the value, $\varphi(x)$, is the well threshold of x , that is, the terminal critical value of f_a , $a = f(x)$, at which the component of x in the sublevel set falls ill. The well threshold has another geometric interpretation. Letting $p : [0, 1] \rightarrow \Phi$ be a path on the manifold after surgery, we consider its composition with the mapping, $f \circ p : [0, 1] \rightarrow \mathbb{R}^2$, and write $\ell(p)$ for the length of the image, $f \circ p[0, 1]$. Taking the infimum over all paths that start at x and end on the boundary, $\partial\Phi$, we get $\text{dist}(x) = \inf_p \ell(p)$, the *distance* of $a = f(x)$ from the relevant portion of the contour. We note that $\text{dist}(x)$ is not necessarily the distance to the nearest point on the contour but rather to the nearest point that affects the wellness of the component of x in the sublevel set of f_a .

WELL FUNCTION LEMMA. Let $f : \mathbb{M} \rightarrow \mathbb{R}^2$ be a generic, smooth mapping and $\varphi : \Phi \rightarrow \mathbb{R}$ its well function. Then $\varphi(x) = \text{dist}(x)$ for every point $x \in \Phi$.

PROOF. Let $a = f(x)$. The point x belongs to the zero set of f_a and its component in the sublevel set falls ill at $\mathbb{M}_{\varphi(x)}(a)$. We write $R = \varphi(x)$ for short. The goal is to prove $R = \text{dist}(x)$. It is easy to see that $R \leq \text{dist}(x)$. By the Stability Theorem for Well Diagrams, we have $|\varphi(x) - \varphi(y)| \leq \|f_a - f_b\|_\infty$, and by the triangle inequality in \mathbb{R}^2 , we have $\|f_a - f_b\|_\infty \leq \|a - b\|_2$, where $b = f(y)$. It follows that $\varphi(y) > 0$ for all

points y with $\|a - f(y)\|_2 < R$. Since φ is zero at the boundary, this implies that all points of $\partial\Phi$ are at Euclidean distance at least R from a .

The more difficult direction is to prove $\text{dist}(x) \leq R$. To get a contradiction, we assume $R < \text{dist}(x)$. Let $q : [0, 1] \rightarrow \Phi$ be a path starting at $q(0) = x$ with length $\ell(q) = R$, and let $y = q(1)$ be its endpoint. It belongs to the component C of $\mathbb{M}_R(a)$ that contains x . Since $\varphi(y) > 0$, there is a positive radius δ such that the component, C' , of $\mathbb{M}_\delta(b)$ that contains y is well, that is, the degree of f restricted to C' is non-zero. Since C and C' overlap, their degrees are the same and we can form the union to get a patch, $C \cup C'$, that has the same degree still. We do the same for all points y reachable from x by paths of length R , choosing $\delta > 0$ smaller than the minimum well threshold of any of these points. The result is a component C'' of $\mathbb{M}_{R+\delta}(a)$ that contains C and the restriction of f to C'' has the same degree as the restriction to C . Hence, C'' is well, contradicting the choice of R as the well value of x . \square

Similarity of well functions. We have one more hurdle to clear, namely showing that the well functions for similar mappings are similar. Let $\varphi : \Phi \rightarrow \mathbb{R}$ and $\gamma : \Gamma \rightarrow \mathbb{R}$ be the well functions of the mappings $f, g : \mathbb{M} \rightarrow \mathbb{R}^2$. We say the *difference* between them is at most r , denoted as $\|\varphi - \gamma\|_\infty \leq r$, if there are subspaces $\Phi_0 \subseteq \Phi$ and $\Gamma_0 \subseteq \Gamma$ that contain all points with well threshold r or larger and a bijection $\iota : \Phi_0 \rightarrow \Gamma_0$ such that $f(x) = g \circ \iota(x)$ for every $x \in \Phi_0$ and $g(y) = f \circ \iota^{-1}(y)$ for every $y \in \Gamma_0$. We derive an upper bound on the difference between the two well functions.

HOMOTOPY LEMMA. Let $f, g : \mathbb{M} \rightarrow \mathbb{R}^2$ be two generic, smooth mappings with corresponding well functions $\varphi : \Phi \rightarrow \mathbb{R}$ and $\gamma : \Gamma \rightarrow \mathbb{R}$. Then the difference between the two well functions is $\|\varphi - \gamma\|_\infty \leq \max_{x \in \mathbb{M}} \|f(x) - g(x)\|_2$.

PROOF. We use the straight-line homotopy between f and g defined by $f_t(x) = (1 - t)f(x) + tg(x)$. All f_t are smooth but not necessarily generic. Nevertheless, the well diagram is defined for each distance function $(f_t)_a$. The Stability Theorem for Well Diagrams holds also for non-generic functions, implying that the points in these diagram vary continuously with a and t . Specifically, the bottleneck distance between the diagrams of $(f_t)_a$ and $(f_{t'})_a$ is bounded from above by $|t - t'| \varepsilon$, where ε is the maximum Euclidean distance between corresponding images, as before.

To relate φ with γ , we pick a point $\varphi(x)$ in the well diagram of $f_a = (f_0)_a$. Initializing the construction of a function $\alpha : [0, 1] \rightarrow \mathbb{R}$, we set $\alpha(0) = \varphi(x)$. Increasing t , we continuously extend α until we either reach $t = 1$ or α vanishes. Whenever we reach $t = 1$, we get a point $y \in \Gamma$ with $\alpha(1) = \gamma(y)$. Because the slope of α is between $\pm\varepsilon$, we have $|\varphi(x) - \gamma(y)| \leq \varepsilon$. Collecting all pairs (x, y) generated by this process, we get the bijection $\iota : \Phi_0 \rightarrow \Gamma_0$ required by the claim. We get $\varphi(x) < \varepsilon$ for all $x \in \Phi - \Phi_0$ because α vanishes before reaching $t = 1$. The construction of the functions α can also be done in the other direction, starting at $t = 1$. Making sure we get the same pairs, we also get $\gamma(y) < \varepsilon$ for all $y \in \Gamma - \Gamma_0$, as required. \square

Note that the paths connecting points x with y form a homeomorphism between Φ_0 and Γ_0 , unless there are branch points in the graph of the homotopy connecting f and g . In the absence of such branch points, we can substitute a homeomorphism for the

bijection in the definition of difference between well functions and embeddings for the injections in the definition of erosion distance.

Finale. We are now in a position to tie up all ends and finish the proof of the Contour Stability Theorem. Let $\varphi : \Phi \rightarrow \mathbb{R}$ and $\gamma : \Gamma \rightarrow \mathbb{R}$ be the well functions of the mappings f and g . By the Well Function Lemma, eroding the 2-manifolds with boundary is the same as taking superlevel sets of the well functions, $\Phi^{-r} = \varphi^{-1}[r, \infty)$ and $\Gamma^{-r} = \gamma^{-1}[r, \infty)$. By the Homotopy Lemma, $\|\varphi - \gamma\|_\infty \leq \varepsilon$. We recall that this means there is a bijection, $\iota : \Phi_0 \rightarrow \Gamma_0$, that is compatible with the two mappings. Here, $\Phi_0 \subseteq \Phi$ and $\Gamma_0 \subseteq \Gamma$ contain all points with well threshold ε or larger, that is,

$$\begin{aligned}\Phi^{-\varepsilon} &= \varphi^{-1}[\varepsilon, \infty) \subseteq \Phi_0; \\ \Gamma^{-\varepsilon} &= \gamma^{-1}[\varepsilon, \infty) \subseteq \Gamma_0.\end{aligned}$$

Restricting the bijection to the superlevel set of φ , we get the injection $\iota_f : \Phi^{-\varepsilon} \rightarrow \Gamma$ defined by $\iota_f(x) = \iota(x)$. Symmetrically, restricting it to the superlevel set of γ , we get the injection $\iota_g : \Gamma^{-\varepsilon} \rightarrow \Phi$ defined by $\iota_g(y) = \iota^{-1}(y)$. By construction, $f(x) = g \circ \iota_f(x)$ for every $x \in \Phi^{-\varepsilon}$ and $g(y) = g \circ \iota_g(y)$ for every $y \in \Gamma^{-\varepsilon}$. It follows that the erosion distance between the 2-manifolds with boundary is $E(\Phi, \Gamma) \leq \varepsilon$, which completes the proof of the Contour Stability Theorem.

6 Discussion

An immediate application of our result is to the artistic representation of shapes using contours. Instead of the entire contour, or perhaps the entire visible contour, we advocate drawing only the portion that remains after a small erosion. A similar strategy may be used to improve the efficacy of shape matching methods that work by comparing contours [18].

The current statement of the Contour Stability Theorem is based on injections in the definition of the erosion distance. It would be nice to replace them by embeddings, but possible branch points in the homotopy as constructed in the proof of the Homotopy Lemma would contradict their existence. Can we find an explicit example in which at least one branch point occurs? Can we substitute piecewise embeddings for the injections?

References

1. P. K. AGARWAL, H. EDELSBRUNNER, J. HARER AND Y. WANG. Extreme elevation on a 2-manifold. *Discrete Comput. Geom.* **36** (2006), 553–572.
2. V. I. ARNOLD. *Catastrophe Theory*. Third edition, Springer-Verlag, Berlin, Germany, 1992.
3. H. BLUM. A transformation for extracting new descriptors of shape. In *Models for the Perception of Speech and Visual Form*, W. Wathen-Dunn (ed.), MIT Press, Cambridge, Massachusetts, 362–380, 1967.
4. G. CARLSSON AND V. DE SILVA. Zigzag persistence. Manuscript, Dept. Mathematics, Stanford Univ., Stanford, California, 2008.

5. D. COHEN-STEINER, H. EDELSBRUNNER AND J. HARER. Stability of persistence diagrams. *Discrete Comput. Geom.* **37** (2007), 103–120.
6. F. COLE, A. GOLOVINSKIY, A. LIMPAECHER, H. S. BARROS, A. FINKELSTEIN, T. FUNKHOUSER AND S. RUSINKIEWICZ. Where do people draw lines? SIGGRAPH Conf. Proc., *ACM Trans. Graphics* **27** (2008), 1–11.
7. D. DECARLO, A. FINKELSTEIN, S. RUSINKIEWICZ AND A. SANTELLA. Suggestive contours for conveying shape. *AMC Trans. Graph.* **22** (2003), 848–855.
8. F. DUGUET AND G. DRETTAKIS. Robust epsilon visibility. SIGGRAPH Conf. Proc., *ACM Trans. Graphics* **21** (2002), 567–575.
9. H. EDELSBRUNNER AND J. HARER. Jacobi sets of multiple Morse functions. In F. Cucker, R. Devore and P. Olver (eds.), *Foundations of Computational Mathematics, Minneapolis 2002*, 35–57, Cambridge University Press, 2004.
10. H. EDELSBRUNNER, D. LETSCHER AND A. ZOMORODIAN. Topological persistence and simplification. *Discrete Comput. Geom.* **28** (2002), 511–533.
11. H. EDELSBRUNNER, D. MOROZOV AND A. PATEL. Quantifying transversality by measuring the robustness of intersections. Manuscript, Dept. Comput. Sci., Duke Univ., Durham, North Carolina, 2009.
12. D. J. FUTUYAMA. *Evolutionary Biology*. Third edition, Sinauer Associates, 1998.
13. M. GLISSE AND S. LAZARD. An upper bound on the average size of silhouettes. *Discrete Comput. Geom.* **40** (2008), 241–257.
14. M. GOLUBITSKY AND V. GUILLEMIN. *Stable Mappings and Their Singularities*. Springer-Verlag, New York, 1973.
15. T. ISENBERG, B. FREUDENBERG, N. HALPER, S. SCHLECHTWEG AND T. STROTHOTTE. A developer’s guide to silhouette algorithms for polygonal models. *IEEE Comput. Graph. Appl.* **23** (2003), 28–37.
16. L. KETTNER AND E. WELZL. Contour edge analysis for polyhedron projections. In *Geometric Modeling: Theory and Practice*, 379–394, eds. W. Straßer, R. Klein and R. Rau, Springer-Verlag, 1996.
17. J. J. KOENDERINK. What does the occluding contour tell us about solid shape? *Perception* **13** (1984), 321–330.
18. P. MIN, J. CHEN AND T. FUNKHOUSER. A 2D sketch interface for a 3D model search engine. SIGGRAPH Technical Sketches (2002), 138.
19. P. V. SANDER, X. GU, S. J. GORTLER, H. HOPPE AND J. SNYDER. Silhouette clipping. SIGGRAPH Conf. Proc., *Computer Graphics* (2000), 327–334.
20. H. WHITNEY. On singularities of mappings of Euclidean space. I. Mappings of the plane to the plane. *Ann. Math.* **62** (1955), 374–410.