Available online at www.sciencedirect.com
ScienceDirect
Theoretical Computer Science 365 (2006) 184–198
Theoretical Computer Science
www.elsevier.com/locate/tcs

Computational topology for isotopic surface reconstruction

K. Abe[a,1], J. Bisceglio[b,2], D.R. Ferguson[c], T.J. Peters[d,*,3], A.C. Russell[e,4], T. Sakkalis[f,g,5]

[a] *Department of Mathematics, University of Connecticut, Storrs, CT 06269-3009, USA*
[b] *Department of Computer Science, University of Connecticut, Storrs, CT 06269-3009, USA*
[c] *DRF Associates, Seattle, WA, USA*
[d] *Department of Computer Science and Engineering and Department of Mathematics, University of Connecticut, Storrs, CT 06269-3155, USA*
[e] *Department of Computer Science and Engineering University of Connecticut, Storrs, CT 06269-3155, USA*
[f] *Agricultural University of Athens, Athens 118 55, Greece*
[g] *Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

Abstract

New computational topology techniques are presented for surface reconstruction of 2-manifolds with boundary, while rigorous proofs have previously been limited to surfaces without boundary. This is done by an intermediate construction of the *envelope* (as defined herein) of the original surface. For any compact C^2-manifold M embedded in $\mathbf{R}^3$, it is shown that its envelope is $C^{1,1}$. Then it is shown that there exists a piecewise linear (PL) subset of the reconstruction of the envelope that is ambient isotopic to M, whenever M is orientable. The emphasis of this paper is upon the formal mathematical proofs needed for these extensions. (Practical application examples have already been published in a companion paper.) Possible extensions to non-orientable manifolds are also discussed. The mathematical exposition relies heavily on known techniques from differential geometry and topology, but the specific new proofs are intended to be sufficiently specialized to prompt further algorithmic discoveries.

Keywords: Ambient isotopy; Computational topology; Computer graphics; Surface approximation; Topology methods for shape understanding and visualization

* Corresponding author.
E-mail addresses: abe@math.uconn.edu (K. Abe), david.r.ferguson@comcast.net (D.R. Ferguson), tpeters@cse.uconn.edu (T.J. Peters), acr@cse.uconn.edu (A.C. Russell), stp@aua.gr (T. Sakkalis).

[1] Partial funding for K. Abe was from NSF Grants CCF 0429477 and CCR 0226504. All statements in this publication are the responsibility of the authors, *not* of these funding sources.

[2] Partial funding for J. Bisceglio was from NSF Grant DMS-0138098 and from a 2005 IBM Faculty Award. All statements in this publication are the responsibility of the authors, *not* of these funding sources.

[3] Partial funding for T.J. Peters was from NSF Grants CCF 0429477, DMS 0138098, CCR 0226504 and from a 2005 IBM Faculty Award.

[4] Partial funding for A.C. Russell was from NSF Grants CCF 0429477 and CCR 0226504.

[5] Partial funding for T. Sakkalis was obtained from NSF Grants DMS-0138098, CCR 0231511, CCR 0226504 and from the Kawasaki chair endowment at MIT.

0304-3975/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.tcs.2006.07.062

Fig. 1. Stadium curve.

1. Introduction and motivation

Several recent approaches to topology-preserving surface approximation have been restricted to C^2 2-manifolds without boundary [2,4,5,9,22]. Generalizations are presented here to:

- those 2-manifolds with boundary which are C^2, and
- those 2-manifolds without boundary which are merely $C^{1,1}$ (see below).

Surface reconstruction is a topic of current interest in computer science and industrial communities. The basic issue is the creation of algorithms for generating an approximating surface from a discrete set of sample points of a surface. For surface reconstruction, the sampled points should also *explicitly* be points of a piecewise linear (PL) output surface. More generally, surface approximations need not be PL and the sampled points need not lie on the output surface. In both cases, it is desirable to have upper bounds on the approximation and to have guarantees of topological equivalence between the original and output surfaces.

The theory presented here is a fundamental step to creating algorithms for reconstruction of surfaces with boundary [9], *with* provable topological characteristics and error bounds for the output surface. A recently published paper by these authors also shows applications of the theory presented here [1]. The main theorem [6] is now stated to motivate the remainder of the paper, with the proof appearing in Section 6.

Theorem 1.1. *If M is a compact, C^2, orientable 2-manifold with boundary, with M embedded in $\mathbf{R}^3$, then there exists a PL ambient isotopic approximation of M, which can be made arbitrarily close to M.*

Our development will rely upon hypotheses of $C^{1,1}$-continuity, so that definition is provided below, whereas other standard terminology from differential geometry and topology are provided for the reader's reference in the Appendix [13,14]. This perspective on weaker differentiability assumptions leads to three new valuable insights for the development and use of surface reconstruction algorithms:

- The proofs on manifolds with boundary hold promise for provably correct algorithms on these difficult cases. (Experimental algorithmic results are discussed elsewhere [1].)
- The generalizations presented for this larger admissibility class of $C^{1,1}$-surfaces can be used for the construction of data filters to preclude inadmissable input.
- An example on a Möbius strip shows potentional extensions to non-orientable surfaces, versus typical assumptions about orientability within the computer science literature.

Definition 1.1. A real-valued function f, defined in an open subset U of R^3, is said to be a $C^{1,1}$ function if its gradient ∇f is Lipschitz continuous in U.

Definition 1.2. An embedded manifold (M, f) is $C^{1,1}$ if $f(M)$, is locally given as the graph of a bivariate, real-valued $C^{1,1}$ function.

The well-known stadium curves, as illustrated in Fig. 1 are $C^{1,1}$ but not C^2 and these easily generalize to surface examples.

[6] The theorem is stated explicitly for manifolds with boundary, as the case of manifolds without boundary was previously proven [5]. However, the proof given here also works for manifolds without boundary, so the theorem given here can be understood to be inclusive of both those cases.

2. Related work

There are several recent publications [2,4,5,9] with an emphasis upon topological guarantees for surface reconstruction. This paper presents significant theoretical extensions beyond that cited literature, as noted in the previous section. Furthermore, examples showing the power of these theoretical extensions are published elsewhere by these authors [1] and the interested reader is referred there for further details. The theoretical concerns in providing topological guarantees for surface approximations near boundaries have been presented in the literature [3,9,12] within the context of approximants created during surface reconstruction.

The value in preferring ambient isotopy for topological equivalence [5,22] versus the more traditional equivalence by homeomorphism [24] has previously been presented [5,22] and the reader is referred to those papers for formal definitions.

Basic notions from differential topology and geometry are summarized in the Appendix [13,14] and readers familiar with this material may use it primarily as a reference for the notation that appears in the rest of the paper.

3. Proof overview and definition of the envelope

Remark 3.1. All surfaces are assumed to be compact 2-manifolds embedded within $\mathbf{R}^3$.

An overview of the primary computational topology technique is now given. For a C^2-manifold, M, with boundary, it is shown that a $C^{1,1}$-manifold without boundary can be constructed arbitrarily close to M. By extensions presented here, this $C^{1,1}$-manifold is admissable input to present surface reconstruction algorithms. It is then shown that a mapping from a subset of the reconstruction of the $C^{1,1}$-manifold is ambient isotopic to M. Furthermore, it is shown that the medial axis [5] of the $C^{1,1}$-manifold is equal to M and this aspect is exploited in preliminary experimental examples presented here and in our companion paper [1]. Existing algorithms can also produce approximations to the medial axis, but there remain open issues regarding topological guarantees, numerical properties and acceptable performance of those medial axis approximation algorithms—all of which lie beyond the scope of the current work, but are being considered within the broader research community [2,25].

The purpose of the rest of this section is to define a new surface that can be created from M, which we call the envelope of M. Some properties of the envelope are then proven. These proofs rely upon the use of boundary collars [14] as well as upon an upper bound between M and its envelope, to ensure that the resulting envelope will not be self-intersecting or degenerate. Let M be a surface with boundary. Then we have from the definition:

(1) ∂M is a disjoint union of closed curves $c_1, \dots, c_l$, each of which is diffeomorphic to the unit circle S^1.

(2) Along each c_j, $1 \leqslant j \leqslant l$, we can attach a collar of the form $c_j \times [0, 2\varepsilon_j)$, for some positive number ε_j, so that the topological space $M_j = M \cup (c_j \times [0, 2\varepsilon_j))$ (where M_j is defined under the quotient map that identifies c_j and $c_j \times 0$ in the natural way) is a surface with the same degree of differentiability as M. The surface M_j contains M and M_j now has the previous boundary component in its interior. Thus, successive attachments of collars along all boundary components produce an open surface

$$\tilde{M} = M \cup \left(\bigcup_{1 \leqslant j \leqslant l} (c_j \times [0, 2\varepsilon_j)) \right).$$

$\tilde{M}$ contains M as a proper subset and $\partial \tilde{M} = \emptyset$. Furthermore, we can choose ε_j, $1 \leqslant j \leqslant l$, in such a way that the embedding f of M can be extended to an embedding $\tilde{f}$ of $\tilde{M}$. This means the pair $(\tilde{M}, \tilde{f})$ is a surface in $\mathbf{R}^3$ which extends the original surface (M, f).

For technical reasons within the proofs, we introduce a new surface $\hat{M}$ with boundary $\partial \hat{M}$ given by $\hat{M} = \tilde{M} - \bigcup_{j=1}^{l} (c_j \times (\varepsilon_j, 2\varepsilon_j))$. We note here that the minimal positive critical values of the global energy function G defined in $\hat{M}$ (see Appendix) are less than or equal to that in M.

With respect to the induced metric in $\tilde{M}$ from $\mathbf{R}^3$, consider a unit normal field ξ to $\tilde{M}$. The shape operator A_ξ of $\tilde{M}$ is given as the tangential component of the directional derivative of ξ; namely, $\tilde{f}_*(A_\xi(X)) = -D_X \xi$, which is the directional derivative of ξ in the x-direction. The operator D is also called the covariant derivative in differential geometry (or often the standard Riemannian connection).

Let $\hat{c}$ denote the smallest positive critical value of $\hat{G}$, the natural extension of G to $\hat{M} \times \hat{M}$, where $\hat{c}$ is less than the minimal critical value for G. Also, denote by $\kappa = \max_{x \in M} \{|K_1(x)|, |K_2(x)|\}$, where $K_i(x),\ i = 1, 2$ are the principal curvatures at $x \in M$. Now denote by $\hat{\kappa}$ the number defined to be $\max_{x \in \hat{M}} \{\hat{K}_1(x)|, |\hat{K}_2(x)|\}$, where $\hat{K}_i(x),\ i = 1, 2$ are the principal curvatures at $x \in \hat{M}$. As noted before these are at least continuous in M and $\hat{M}$, respectively. Then $\kappa \leqslant \hat{\kappa}$. Since $\hat{M}$ is compact, the absolute values of these quantities attain the absolute extrema.

Definition 3.1. Set $\hat{\delta} = \frac{1}{2} \min\{\hat{c}, 1/\hat{\kappa}\}$.

Note here that we use the convention $1/\kappa = +\infty$ when $\kappa = 0$ without loss of generality. Also note that it is well known that M is a part of a plane if the principal curvatures are zero everywhere in M. We may then exclude this case since an ambient isotopy of such a set can readily be constructed. Hence, we assume $\hat{\delta}$ to be a finite positive number.

We introduce a compact closed surface called the r-envelope of M as follows. Let $c_i,\ 1 \leqslant i \leqslant n$ be the boundary curves of M. We first define a surface $P_r(c_i)$ about $c_i,\ 1 \leqslant i \leqslant n$ (each such surface is called a pipe surface [17]). A specific parametrization of these surfaces is given for later use.

Let $c = c(t),\ t \in [0, l]$ be a regular closed space curve in $\mathbf{R}^3$. Further assume that the curve has no self-intersection and that it is parametrized by its arc length; hence, l is the total arc length of the curve. For a sufficiently small $r > 0$,

$$P_r(s, t) = c(t) + r\xi(t) \cos s + r\eta(t) \sin s, \quad 0 \leqslant t < l,\ 0 \leqslant s < 2\pi$$

gives rise to a closed surface in $\mathbf{R}^3$ parametrized by (s, t), where $\xi(t)$ and $\eta(t)$ form an orthonormal frame normal to the curve. For example, they can be the pair consisting of the normal and binormal of the curve [20]. We have $P_r(s, t) = P_r(c_i)$ when $c = c_i$. One may consider (t, s) as its coordinates (see the remark below). The tangent plane to this surface at (t, s) is spanned by the following two tangent vectors:

$$\frac{\partial}{\partial t} = \frac{\partial(c(t) + r\xi(t) \cos s + r\eta(t) \sin s)}{\partial t} = \frac{\mathrm{d}c(t)}{\mathrm{d}t} + r\frac{\mathrm{d}\xi}{\mathrm{d}t} \cos s + r\frac{\mathrm{d}\eta}{\mathrm{d}t} \sin s,$$
$$\frac{\partial}{\partial s} = \frac{(\partial c(t) + r\xi(t) \cos s + r\eta(t) \sin s)}{\partial s} = -r\xi(t) \sin s + r\eta(t) \cos s.$$

One can readily see from the above expressions that these tangent vectors are linearly independent for sufficiently small r, hence, the resulting surface is indeed an embedded surface in $\mathbf{R}^3$. The surface $P_r(c)$ for each sufficiently small r is called the r-pipe surface [17]. It is the well-known embedded r circle bundle of the curve. The radial vectors emanating from $c(t)$ are the radial vectors of the circles. Hence, they are given by $r\xi(t) \cos s + r\eta(t) \sin s,\ 0 \leqslant t < l,\ 0 \leqslant s \leqslant 2\pi$. We show that these radial vectors are, indeed normal to the surface at each (t, s). First note the following:

(i) $(\mathrm{d}c(t)/\mathrm{d}t) \cdot \xi = (\mathrm{d}c(t)/\mathrm{d}t) \cdot \eta = 0$, with '$\cdot$' denoting dot product.
(ii) $(\mathrm{d}\xi/\mathrm{d}t) \cdot \xi = (\mathrm{d}\eta/\mathrm{d}t) \cdot \eta = 0$, since ξ and η are unit vectors.

Using (i) and (ii), one can easily compute that the dot products between $\partial/\partial t$ and $\partial/\partial s$ and the radial vectors are 0; hence, the radial vectors will be normal to the r-envelope, as defined below, in Definition 3.2.

It is known [14] that there is a certain positive number δ_c such that the map given by $(s, t, r) \mapsto c(t) + r\xi(t) \cos s + r\eta(t) \sin s,\ 0 \leqslant t < l,\ 0 \leqslant s < 2\pi,\ 0 \leqslant r < \delta$ is an embedding into $\mathbf{R}^3$. This is typically called the *r-tubular neighborhood* and is a subset of the r-envelope, defined below (Definition 3.2).

Let x be a point in ∂M. We may assume that x belongs to a C^2-regular space curve $c_i = c_i(t),\ 0 \leqslant t < l_i$ with $c_i(0) = x$. We may even assume c_i is parametrized by its arc length without loss of generality. This implies $|\mathrm{d}c_i/\mathrm{d}t| \equiv 1$ for all t and that l_i equals the arc length of c_i. Let ξ be a unit normal to M. Denote by $\xi(t)$ and $\eta(t)$ the restriction of ξ to c_i and the unit outward normal at $c_i(t)$, respectively, so chosen that $\mathrm{d}c/\mathrm{d}t, \xi(t), \eta(t)$ form the right-hand system relative to standard orientation in $\mathbf{R}^3$. Here the outward normal means the unit vector that is perpendicular to the plane spanned by $\mathrm{d}c_i/\mathrm{d}t$ and ξ and that points away from M at $c_i(t)$. Since M is C^2, these vectors are at least C^1 along $c_i(t)$.

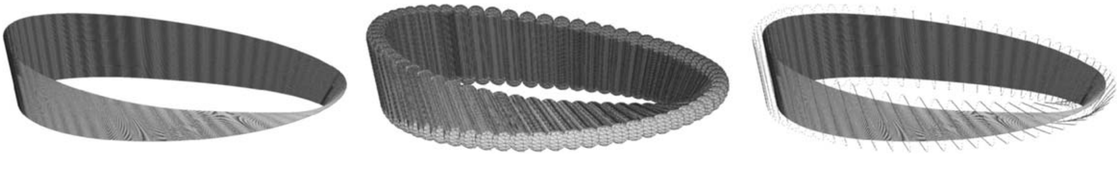

Fig. 2. Möbius strip.

Fig. 3. Comparison of methods: Möbius strip.

For any $r > 0$, define $E_r(M)$ by

$$E_r(M) = \{x \pm r\xi : x \in M\} \cup \{c_i(t) + r\xi(t)\cos s + r\eta(t)\sin s : 0 \leqslant t < l_i,\ 0 \leqslant s < \pi\}.$$

Definition 3.2. $E_r(M)$ is called the r-envelope [7] of M.

Remark 3.2. We note that M has been assumed to be orientable, as this is an important hypothesis for Theorem 6.1, a primary result of this paper. However, the definition of the envelope does not depend on a particular choice of the (local) unit normal to the surface and the envelope construction is purely local in nature. The same (but local) analytic representation of the envelope as stated above will uniquely and globally define the envelope regardless of orientability of the surface. Hence, $E_r(M)$ remains well-defined even if M is non-orientable, while $E_r(M)$ will always be an orientable surface. The orientability of $E_r(M)$ is exploited in the computational experiments presented showing an approximation of a Möbius strip via the medial axis of its envelope, as illustrated in Figs. 2 and 3. This computational result was motivated by our main theorem.

An illustration of this envelope is given in Fig. 2, below. The surface M depicted is a Möbius strip, to emphasize that the definition of $E_r(M)$ is independent of the orientability of M, but similar images will exist for orientable 2-manifolds. The constructive computational process is given across the three sub-images. The left one is just a standard-tessellated

[7] Wolter [25] constructed the envelope of a spline surface parametrized in $\mathbf{R}^3$ by $[0, 1] \times [0, 1]$, although he did not call it an envelope. He states without proof that this envelope is a $C^{1,1}$-surface which is $C^{1,1}$-diffeomorphic to the unit two-dimensional sphere for sufficiently small r. Strictly speaking, our proof is not applicable to his case since $[0, 1] \times [0, 1]$ is not a surface with boundary according to our Definition A.2.

graphics display of a Möbius strip. The middle image is a graphical display of many balls, centered at each vertex of the Möbius strip tessellation, to give a visual representation of the r-envelope, for a radius r. The right shows the Möbius strip and points sampled from its r-envelope.

Note that $E_r(M)$ is not even a topological manifold for some r, but it is readily seen that for a sufficiently small r, $E_r(M)$ is at least C^1 everywhere, except possibly in a finite number of curves where it is at least G^1. We now give an explicit description of those curves for the future use. Set $S_i(r,t) = c_i(t) + r\xi(t),\ |r| < \hat{\delta},\ 0 \leqslant i \leqslant n$, where c_i's are the boundary components and ξ is the unit normal to M along those components. Note that at this point S_i may not be a regular surface, but it is the union of open line segments of length $2\hat{\delta}$ centered at the points in $c_i(t)$. In fact, they are ruled surfaces built on the boundary curves with $\xi(t)$ as the direction of the rulings. The set $E_r(M) \cap S_i(r,t)$ gives rise to a curve in $E_r(M)$ for each fixed r. Denote such a curve by $S_{i,r}$ for each i. In fact, we will show later that $E_r(M)$, for a certain range of r to be specified later, is C^2 everywhere but along the $S_{i,r}$'s where it is at least $C^{1,1}$.

Now set

$$\delta = \min\{\hat{\delta}, \delta(c_i) :\ 1 \leqslant i \leqslant n\}, \tag{1}$$

where $c_i,\ 1 \leqslant i \leqslant n$ is a boundary curve of M and $\delta(c_i)$ is the maximal radius of the regularly embedded pipe surfaces $E_r(c_i)$ [15].

Let r_0 be a sufficiently small positive number so that $E_{r_0}(M)$ is well-defined and C^1 except for along the curves S_{i,r_0} where it is G^1. Define a map

$$F_{r_0} : E_{r_0}(M) \times (-r_0, \hat{\delta} - r_0) \to \mathbf{R}^3$$

by

$$F_{r_0}(x,r) = x + r\mathbf{n}, \quad (x,r) \in E_{r_0}(M) \times (-r_0, \hat{\delta} - r_0), \tag{2}$$

where $\mathbf{n}$ is the unit normal field to $E_{r_0}(M)$ which points away from M at each point of $E_{r_0}(M)$. Such a choice of a normal is possible because of the definition of the envelope.

Lemma 3.1. *$F_{r_0}(x,r)$ is globally injective.*

Proof. First, we clearly see that $F_{r_0}(x,r)$ is a globally injective C^1-diffeomorphism when it is restricted to the pipe surface portions of the envelope by the choice of $\hat{\delta}$. Furthermore, the implicit function theorem yields $E_r(M)$ is a C^1-surface in the neighborhood of the points in the pipe surface portions. For any point $x \in E_r(M)$ given by the expression $\{x \pm r\xi :\ x \in M - \partial M,\ 0 < r < \hat{\delta}\}$, we need somewhat more elaborate and lengthy (but more or less elementary) arguments, for which we only give an outline here to save space. First we enlarge the set to $\{x \pm r\xi :\ x \in \hat{M} - \partial\hat{M},\ 0 < r < \hat{\delta}\}$. Now define a map $F : (\hat{M} - \partial\hat{M}) \times (-\hat{\delta}, \hat{\delta}) \to \mathbf{R}^3$ by

$$F(x,r) = x + rn, \quad r \in (-\hat{\delta}, \hat{\delta}), \tag{3}$$

where $n = n_x$ is a unit normal to $\hat{M}$ at x. Then it is well known that the Jacobian map F_* of F at (x,r) is the symmetric linear map whose eigenvalues are given by $K_i/(1 - rK_i)$ and 1, where $K_i,\ i = 1, 2$ are the principal curvatures of $(\hat{M}, f)$. Consequently, F is non-singular as long as $|r| < \hat{\delta}$. This implies that F is locally a C^1-diffeomorphism since $\hat{M}$ is a C^2 surface. Hence, $F_{r_0}(x,r)$ is locally injective near every $(x,r) \in E_{r_0}(M) \times (-r_0, \hat{\delta} - r_0) - P_\delta$, where $P_\delta = \bigcup_{r<\delta,\ 1 \leqslant i \leqslant n} P_{i,r}(s,t)$, with each $P_{i,r}(s,t)$ being the previously defined set $P_r(s,t)$ specific to the curve c_i. Note that $F_{r_0}(x,r)$ is basically defined by restricting F to this set.

Finally along the $S_i = c_i(t) + r\xi(t),\ |r| < \hat{\delta},\ 0 \leqslant i \leqslant n$, it is not hard to see that the envelope is G^1, i.e. the tangent planes vary continuously, and that $F_{r_0}(x,r)$ is locally injective along S_i's by the definition of the envelope and the local injective property of F_{r_0} off S_i's as described above.

To show that F_{r_0} is globally injective, first note that F_{r_0} is globally injective in $E_{r_0}(M) \times (-r_0, 0]$ by the choice of r_0. Let $\varepsilon_0 = \inf \varepsilon$ such that $F_{r_0}(x,r)$ fails to be globally injective in $E_{r_0}(M) \times (-r_0, \varepsilon)$. It can be shown then that the existence of such an ε_0 less than $\delta - r_0$ presents a contradiction to the choice of δ, using arguments similar to ones previously published [22,23], now applied to $E_{r_0}(M)$ in place of the compact closed surface M. Note that $E_{r_0}(M)$ is a compact closed surface. Although $E_{r_0}(M)$ is not C^2 as assumed in [5,23], the basic arguments still applies to $E_{r_0}(M)$ with slight modifications. □

4. Isotopies of the envelope

Given a point $x \in \mathbf{R}^3$, define a real-valued function ρ_M by $\rho_M(x) =$ the ordinary distance function from x to M. Since M is compact, there is a point $m_x \in M$ such that $\rho_M(x) = |x - m_x|$. Since M is a C^2-surface (with or without boundary), the line joining x and m_x meets M perpendicularly. Thus, m_x is the foot of the perpendicular projection of x onto M. From Lemma 3.1 above, m_x is uniquely determined if x lies in the (connected) component of $\mathbf{R}^3 - E_{\hat{\delta}}(M)$ which contains M. The component is an open neighborhood of M. Denote it by U_δ. This tells us that ρ_M is well-defined in U_δ. We know in general such a ρ_M is Lipschitz continuous.

Theorem 4.1. *The distance function ρ_M is a C^2-function in $U_\delta - M = \bigcup_{0<r<\delta} E_r(M)$ except along a finite number of surfaces S_i, $1 \leqslant i \leqslant n$, where it is $C^{1,1}$. The envelope $E_r(M)$, $0 < r < \delta$ is C^2 everywhere except along the curves*

$$S_{i,r} = c_i(t) \pm r\xi(t), \quad 0 < r < \delta, \ 1 \leqslant i \leqslant n,$$

and it is at least $C^{1,1}$ along those curves.

Proof. Restrict the distance function ρ_M to the following two subsets:

$$v_\delta = \{x \pm rn(x) : \ x \in M - \partial M, \ 0 < r < \delta\},$$

where $n(x)$ is the normal to M at x and $\xi = n$ along ∂M and

$$B_\delta = \{c_i(t) + r\xi(t)\cos s + r\eta(t)\sin s : \ 0 \leqslant t < l_i, \ 0 < s < \pi, \ r < \delta\}.$$

We first show that the distance function ρ defined in these sets are C^2-functions. $F(x, r) = x + rn(x)$, $|r| < \delta$ is locally diffeomorphic at $x \in M - \partial M$ by the choice of δ. It is not hard to see that this diffeomorphism is at least a C^1-diffeomorphism, since the Jacobian map of F is locally given in terms of the shape operator of the r level set $F(x, r)$, where $r \in [0, \delta)$ is fixed to be a constant. Note that the shape operators (or their eigenvalues) are at least continuous [14]. Thus, we may consider F as giving a C^1-local coordinate chart about every point in v_δ. With this coordinate system, it is easy to see that the gradient field $\nabla\rho$ of the distance function ρ is the unit tangential field to the normal rays emanating from M. The normal rays are generated by the normal field n to M and n is at least C^1, since M is assumed to be C^2. Hence, the tangential field is C^1. This implies that the gradient field $\nabla\rho$ is a C^1-field; consequently, ρ is a C^2-function in $v_\delta = \{x \pm rn(x) : \ x \in M - \partial M, \ 0 < r < \delta\}$. Applying the implicit function theorem, the level sets of the distance function are also C^2 in v_δ. Similarly, we see that the gradient field of the distance function in B_δ is the unit C^1-field generated by the radial rays emanating from the boundary of M. This is an easy consequence of our choice of δ [10,15]. One can, in fact, show that the map $F : (0, \delta) \times \mathbf{R}^2(s, t) \to \mathbf{R}^3(x, y, z)$ defined by

$$F(r, s, t) = \{c_i(t) + r\xi(t)\cos s + r\eta(t)\sin s : \ 0 \leqslant t < l_i, \ 0 \leqslant s \leqslant 2\pi, \ r < \delta\} \quad (4)$$

is at least a C^1-diffeomorphism. This, in turn, yields that the gradient field $\nabla\rho$ of the distance function $\rho(F(r, s, t)) = r$ coincides with the radial unit normal which is defined to be the field of the unit tangent vectors to the radial rays that emanate from each point of c_i into the normal directions to the curve c_i at the point; hence, the desired result. Once again, one can show that the radial normal field is at least C^1. Thus, the distance function ρ_M is a C^2-function in B_δ. The implicit function theorem again yields the desired result that the level surfaces of the distance function ρ_M are C^2-surfaces except at $r = 0$, where it degenerates to be the boundary curves.

We now construct a specific C^1-local coordinate chart $(\tilde{U}_m, \psi_m)$ in $\mathbf{R}^3$ about every point m in the surface $S_i(r, t) = c_i(t) + r\xi(t)$, $0 < r < \delta$, $0 \leqslant t < l_i$. Let $\eta_i(r, t)$ be the outward unit normal field to the $S_i(r, t)$. Then $\eta_i(r, t)$ is a local C^1-vector field along $S_i(r, t)$ and it is tangent to $E_r(M)$. Note that the surfaces S_i's are actually at least C^1-surfaces. This can be verified by realizing that these surfaces occur in the interior of the solid pipes over the boundary components, or can be regarded as surfaces in $\{x \pm rn(x) : \ x \in \hat{M} - \partial\hat{M}, \ 0 < r < \delta\}$, where $n(x)$ is the normal to $\hat{M}$ at x and $\xi = n$ along $\partial\hat{M}$. Define a new vector field $\tilde{\eta}_i$ along $S_i(r,t)$ by $\tilde{\eta}_i(r,t) = r\eta(r,t)$. $\tilde{\eta}_i(r,t)$ is also a C^1-vector field along the surface, since r is clearly a C^1-function there. Extend $\tilde{\eta}_i(r,t)$ to a non-zero C^1-vector field in a neighborhood V_m of m and denote it by the same letter $\tilde{\eta}$ for convenience. Then $\tilde{\eta}$ can be regarded as a C^1-map from $\mathbf{R}^1(t) \times V_m \subset R^4(t,u,v,w)$ into $\mathbf{R}^3(u,v,w)$ by setting $\tilde{\eta}(x) = (\tilde{\eta}_1(x), \tilde{\eta}_2(x), \tilde{\eta}_3(x))$. Consider the system of ordinary differential equations

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = \tilde{\eta}_i, \quad 1 \leqslant i \leqslant 3. \tag{5}$$

By the existence and uniqueness theorem for ordinary differential equations [7] there is a unique solution $x(t) = (x_1(t), x_2(t), x_3(t))$ to this system for a given initial condition in a sufficiently small neighborhood U_m of m, satisfying $x(0) = x_0$, $(\mathrm{d}x/\mathrm{d}t)(0) = \tilde{\eta}(x_0)$. The theorem also states that the local flow $\varphi : (-t_0, t_0) \times U_m \to V_m$ defined by the solutions $\varphi_t(x^0) = x(t; x^0)$ is a C^1-map for a sufficiently small $t_0 > 0$. We choose the set of initial conditions to be the pair $(x, \tilde{\eta}(x))$, $x \in S_i \cap U_m$ and restrict the above map to $(-t_0, t_0) \times S_i \cap U_m$. Note that $S_i \cap U_m$ has a C^1-coordinate system (t, r) induced from the C^1-diffeomorphism (6) above by setting $s = 0$. It is easy to see that this restricted map has a non-degenerate Jacobian map at $(0, m)$. Hence, by the inverse function theorem, this restriction map is a C^1-diffeomorphism in a small neighborhood of m. Denote the diffeomorphism by $\tilde{\psi}_m$ and the neighborhood by $\tilde{U}_m$. The resulting C^1-local coordinate system of the pair $(\tilde{U}_m, \tilde{\psi}_m)$ is denoted by (u, v, w) with $(0, 0, 0)$ representing m. Note here that u, v represent t, $r - r_0$ in $S_i \cap U_m$ and that w is the parameter along the solution curves emanating from $S_i \cap U_m$. Also note that r_0 above corresponds to the radius of the pipe surface that contains m.

Now define ψ_m by

$$\psi_m(u,v,w) = \begin{cases} \tilde{\psi}(u,v,w) & \text{if } w \in (-t_0, 0), \\ c_i(u) + (r_0 + v)\xi(c_i(u))\cos w + (r_0 + v)\eta(c_i(u))\sin w & \text{if } w \in [0, t_0). \end{cases} \tag{6}$$

The function $\psi_m(u,v,w)$ is clearly C^1 except possibly along $w = 0$. The partials $\partial\psi_m/\partial u$, $\partial\psi_m/\partial v$ are continuous even along the surface defined by $w = 0$, hence, they are continuous everywhere. We need to show that $\partial\psi_m/\partial w$ is also continuous along $w = 0$. The partial $\partial\psi_m/\partial w$ is given by tangent vectors of the solutions to the above system of differential equations when $w \leqslant 0$ and it converges to $\hat{\eta}$ as the points approach the surface $w = 0$ from the negative side of w. On the other hand, $\partial\psi_m/\partial w$ is given by the $\partial F/\partial s$ on the positive side of w. The partial $\partial F/\partial s$ converges to $\hat{\eta}$ as $w \to 0$ from right. This, together with the triangle inequality, can be used to show that $\partial\psi_m/\partial w$ is continuous at the points in the surface $w = 0$. Hence, all first partials are continuous in the neighborhood of m. This implies that the map ψ_m is a C^1-map [21, Theorem 9.16]; consequently, ψ_m gives rise to a C^1-coordinate system about m. We are ready to show that the distance function ρ is $C^{1,1}$ along the surfaces S_i, $1 \leqslant i \leqslant n$. We already know that the distance function is C^2 off the surfaces S_i, $1 \leqslant i \leqslant n$. As before, let m be a point in one of S_i, $1 \leqslant i \leqslant n$. Denote by (x, y, z) the standard rectangular coordinates of $\mathbf{R}^3$. Without loss of generality, we may assume that $(0, 0, 0)$ in these coordinates represents m. As we have seen, the gradient field $\nabla\rho$ of ρ is given as the unit tangential field of the normal rays everywhere off the surfaces S_i, $1 \leqslant i \leqslant n$. Since the coordinate transformation between two coordinate systems (x, y, z) and (u, v, w) around m is a C^1-transformation, the induced Jacobian transformation is continuous. From the particular choice of the coordinate system (u, v, w), we see that $\nabla\rho$ is continuous and it, indeed, is the unit tangential field to the normal radial ray emanating from the points in M. By the chain rule, we see that $\nabla\rho$ in terms of (x, y, z) is given as a continuous function of (u, v, w) off the surfaces S_i, $1 \leqslant i \leqslant n$. Since the coordinate transformation between them is a C^1-diffeomorphism, $\nabla\rho$ in (x, y, z), as (x, y, z) approaches points in S_i, $1 \leqslant i \leqslant n$, must converge to the image of $\nabla\rho$ in terms of (u, v, w) under the Jacobian transformation. Since $\nabla\rho$ in the (x, y, z) coordinates is the unit tangential field to the normal radial ray off the surfaces S_i, it converges to the unit tangential field of the normal rays emanating from the boundary curves c_i, $0 \leqslant i \leqslant n$. Thus, the unit tangential field to the normal radial rays must be a gradient field even along the surfaces S_i, $1 \leqslant i \leqslant n$. Consequently, $\nabla\rho$ is C^1 off the surfaces S_i, $1 \leqslant i \leqslant n$, and continuous along those surfaces. We will see that $\nabla\rho$ is Lipschitz continuous along them. To this end, let B_m be a sufficiently small open ball in $\mathbf{R}^3(x, y, z)$ centered at a point m in one of the surfaces S_i, $1 \leqslant i \leqslant n$, say, S_i. We can assume that S_i divides U into two subsets with the common boundary $B_m \cap S_i$. We can also assume that for any $p, q \in B_m$ which belong to the same side of the surface the Lipschitz condition $|(\partial\rho/\partial x)(q) - (\partial\rho/\partial x)(p)| < k|q - p|$ holds. This can be seen as follows. Since p, q belong to the same side of S_i, p, q belong to an open set where ρ is a C^2-function as seen before and $\nabla\rho$ is C^1, hence, Lipschitz. The same observation holds for the other two partials. Now suppose that p, q belong to opposite sides of the surface in B_m. Join p, q by the line segment between them. Since B_m is convex, the entire line segment belongs to B_m. The line segment meets S_i at a point b in B_m. Then the triangle inequality yields

$$\left|\frac{\partial\rho}{\partial x}(q)-\frac{\partial\rho}{\partial x}(p)\right| \leqslant \left|\frac{\partial\rho}{\partial x}(q)-\frac{\partial\rho}{\partial x}(b)\right| + \left|\frac{\partial\rho}{\partial x}(b)-\frac{\partial\rho}{\partial x}(p)\right| < k|q-b|+k|b-p| = k|q-p|.$$

The same proof also works for the other partials. This implies that $\nabla\rho$ is (locally) Lipschitz continuous along the surfaces $S_i,\ 1\leqslant i\leqslant n$; hence, ρ is $C^{1,1}$ there. In particular, applying the implicit function theorem to the distance function, one gets that each level surface is C^2 off $S_i,\ 1\leqslant i\leqslant n$ and $C^{1,1}$ along $S_i,\ 1\leqslant i\leqslant n$. □

With δ defined as in Eq. (1), we have the following corollary.

Corollary 4.2. *The envelope $E_r(M),\ \delta > r > 0$ is the r level surface of the distance function ρ. Furthermore, $E_r(M),\ \delta > r > 0$ form an ambient isotopic family.*

Proof. The first statement is clear from Theorem 4.1. For the second statement, let $0 < r_1 < r_2 < \hat{\delta}$ be any two levels. The gradient field of ρ is given by the unit normal field n. Let ε be a sufficiently small positive number such that $0 < r_1-\varepsilon < r_1 < r_2 < r_2+\varepsilon < \hat{\delta}$ holds. Let f be a positive C^∞ real-valued function satisfying

$$f(r) = \begin{cases} 1 & \text{if } r_1\leqslant r\leqslant r_2, \\ 0 & \text{if } r\leqslant r_1-\varepsilon \text{ or } r\geqslant r_2+\varepsilon. \end{cases} \tag{7}$$

Denote a new vector field $\tilde{n}(r,x)$ in $U_{\tilde{\delta}}$ is defined by $\tilde{n}(x) = f(r)n(x),\ \forall x \in U_{\tilde{\delta}}$. Then $\tilde{n}$ gives rise to a Lipschitz continuous vector field in $\mathbf{R}^3$ with compact support. It generates a one parameter family of diffeomorphisms of $\mathbf{R}^3$ which deforms E_{r_1} onto E_{r_2} [16]. □

Corollary 4.3. *Let M be a compact C^2 surface in $\mathbf{R}^3$. Denote by ∂M its boundary, which could be empty. Denote by M_r the r-offset surface of M. Then for all r, $|r| < \hat{\delta}$, the M_r's are mutually ambient isotopic and the isotopy is obtained through the flow generated by the normal field n to M.*

Proof. If M has no boundary, Corollary 4.3 is proven in [5]. Otherwise, consider $\hat{M}$ introduced earlier, $\hat{M}$ is a C^2-compact surface with boundary $\partial\hat{M}$. The existence of a tubular neighborhood for such a surface tells that there is a sufficiently small $r_0 > 0$ such that all $|r| < r_0$ offset surfaces are ambient isotopic to each other and the ambient isotopy is obtained by the normal flow. This can be seen as follows. Since $\hat{M}$ is compact there is a sufficiently small $r_0 > 0$ such that $F : \hat{M}\times[-2r_0, 2r_0] \to \mathbf{R}^3$ defined by $F(x,r) = x + rn_x,\ |r| < 2r_0$ is an injective diffeomorphism, where n_x is a fixed unit normal field to $\hat{M}$. Both $F(\hat{M}\times[-2r_0, 2r_0])$ and $F(\tilde{M}\times[-r_0, r_0])$ are compact in $\mathbf{R}^3$ and $F(\hat{M}\times[-2r_0, 2r_0])$ contains $F(\tilde{M}\times[-r_0, r_0])$ as a proper subset. It is well known then that there is a C^∞-function $f : \mathbf{R}^3 \to [0, 1]$ such that f is identically 0 outside $F(\hat{M}\times[-2r_0, 2r_0])$ and f is identically 1 inside $F(\tilde{M}\times[-r_0, r_0])$. Let n be the unit tangent vector field to the normal field in $F(\hat{M}\times[-2r_0, 2r_0])$. Then $f\cdot n$ gives rise to a C^1-vector field in $\mathbf{R}^3$ with a compact support. This vector field creates a flow which is identical to the normal flow in $F(\tilde{M}\times[-r_0, r_0])$. Furthermore, the one-parameter family of diffeomorphisms generates the desired ambient isotopy. Now combining this ambient isotopy with the ambient isotopy given in Corollary 4.2 yields the desired ambient isotopy. □

5. Minimum positive distance from the medial axis and a folk lemma

This section presents a lemma which may be of general interest regarding the relation between a surface and its medial axis [5]. It extends the known proofs for the existence of a positive minimum distance between the surface and its medial axis to surfaces which are $C^{1,1}$, as defined, below.

One of the basic consequences from the definition is that M can be locally represented as the "graph" of a real-valued function of two variables. In particular, we can assume that for a given point $m \in M$, there is an open neighborhood $U(x,y,z)$ of 0 in $\mathbf{R}^3$ such that $m = 0$ and the graph of a function $z = f(x,y)$ with $\nabla f(0,0) = 0$ represents M in U, where ∇f denotes the gradient of f.

Lemma 5.1. *For a compact, $C^{1,1}$-manifold M, there exists a positive minimum distance between M and its medial axis.*

Proof. First note that

$$|\nabla f(x, y) - \nabla f(0, 0)| = |\nabla f(x, y)| \leqslant k|(x, y) - (0, 0)| = k|(x, y)|. \tag{8}$$

Hence, along any line given by $ax + by = 0$, or $(t, -a/bt)$, $-\delta < t < \delta$ for a small δ,

$$f(t, (-a/b)t) = \int_\alpha \nabla f(t, -a/bt) \leqslant \int_\alpha |\nabla f(t, -a/bt)| \leqslant k \int_\alpha \sqrt{1 + (a/b)^2} t, \tag{9}$$

where $\alpha = \alpha(t)$ is the space curve given by $\alpha(t) = (t, -a/bt, f(t, -a/bt))$ and the first equality follows from the Fundamental Theorem of Line Integrals. On the other hand, $k \int_\alpha \sqrt{1 + (a/b)^2} t = (k/2)\sqrt{1 + (a/b)^2} t^2$. This yields that

$$f(x, y) = f(t, -a/bt) \leqslant (k/2)\sqrt{1 + (a/b)^2} t^2$$

for all a, b. Note if $b = 0$, just parametrize the y-axis in t. Since

$$(1/2)\sqrt{1 + (a/b)^2} \leqslant 1 + (a/b)^2,$$

the last inequality expressed in terms of x, y gives

$$f(x, y) \leqslant k(x^2 + y^2), \tag{10}$$

in a small neighborhood of $(0, 0)$.

This shows that the graph, therefore the surface, lies below the paraboloid $z = k(x^2 + y^2)$. It is now clear that the curvature sphere of the paraboloid at (0,0,0) is tangent to the graph $z = f(x, y)$ at (0,0,0) and fits entirely above the graph. The equation of the curvature sphere is given by $x^2 + y^2 + (z - (1/2k))^2 = (1/2k)^2$. Applying this argument at every point in M and using the compactness hypothesis, we get a minimum radius λ of the spheres. Then, similar to previous proofs [22,23] a minimum critical value c is defined. Although these previous proofs assumed that the manifold was C^2, the hypothesis here of $C^{1,1}$ is sufficient to derive this value of c. We then define ρ, as

$$\rho = \min\{\lambda, c\}.$$

Then there are no points in the medial axis of M within any distance less than ρ. □

Remark 5.1. This minimum distance Lemma 5.1 can be directly applied to previously presented theorems about C^2-manifolds [5,23], to extend them to compact, $C^{1,1}$ 2-manifolds without boundary.

For the presentation of the main theorem in the next section, the following lemma provides guidance for the construction of an ambient isotopy having a set of compact support. It is likely a "folk theorem". This lemma provides sufficient conditions so that two different isotopies defined over intersecting sets of compact support can be "pasted" together to yield a single ambient isotopy over the union of the sets of compact support.

Lemma 5.2. *For $n \geqslant 0$, let F be an ambient isotopy defined on $\mathbf{R}^n \times [0, 1]$ onto $\mathbf{R}^n$ so that subsets A and B of $\mathbf{R}^n$ are ambient isotopic under F. Similarly, let G be an ambient isotopy defined on $\mathbf{R}^n \times [0, 1]$ onto $\mathbf{R}^n$ so that subsets C and D of $\mathbf{R}^n$ are ambient isotopic under G. Furthermore, suppose that F has compact support CS(F) and G has compact support CS(G). If each point of $x \in CS(F) \cap CS(G)$ is a fixed point of both F and G, then the function*

$$F \cup G : \mathbf{R}^n \times [0, 1] \to \mathbf{R}^n,$$

defined by

$$F \cup G(x, t) = F(x, t) \quad \forall x \in CS(F) \ \ \forall t \in [0, 1],$$

$$F \cup G(x, t) = G(x, t) \quad \forall x \in CS(G) \ \ \forall t \in [0, 1],$$

and

$$F \cup G(x, t) = x \quad \forall x \in \mathbf{R}^n - (CS(F) \cup CS(G)) \ \ \forall t \in [0, 1]$$

is an ambient isotopy with compact support $CS(F) \cup CS(G)$ *such that* $A \cup C$ *is ambient isotopic to* $B \cup D$ *under* $F \cup G$.

Proof. The proof is elementary and complete details are available elsewhere [18]. □

6. Isotopy of the manifold with boundary

This section presents the main theorem of this paper and its proof. This theorem forms the theoretical basis for the existence of an ambient isotopic PL approximation of a compact orientable surface with boundary. Previously, there were only firm theoretical foundations for creation of ambient isotopic PL approximations of manifolds without boundary. Those proofs relied upon the demonstration of a positive minimum distance between the surface and its medial axis, a condition which remains true for $C^{1,1}$-surfaces by Lemma 5.1.

The construction of an ambient isotopic PL approximation of M proceeds by first creating $E_r(M)$, with $r < \delta$ (as defined in Eq. (1)), so that Theorem 4.1 can be invoked. Furthermore, we assume the availability of a simplicial approximation $K(r)$ ambient isotopic to the envelope, $E_r(M)$, where the homeomorphism between $E_r(M)$ and $K(r)$ has specifically been constructed using the nearest point map from $K(r)$ onto $E_r(M)$, so that each point $x \in K(r)$ is mapped into $E_r(M)$ along a normal vector of $E_r(M)$, as can be done according to Remark 5.1. We are then interested in defining a mapping from M into $K(r)$ whose image is ambient isotopic to M. If M is orientable, a consistent normal direction can be chosen on M. Then, for each $x \in M$, its consistently chosen normal vector $\vec{n}_x$ intersects $K(r)$ in a unique point nearest to M, designated as $\Psi(x)$. The correspondence $x \mapsto \Psi(x)$ gives rise to a homeomorphism between M and $\Psi(M)$ since Ψ has been restricted to a specific unit normal direction n to M. Observe that $\Psi(M)$ is not necessarily a simplicial subcomplex of $K(r)$ because the image of ∂M under Ψ will not necessarily be PL. However, it will be possible to obtain an ambient isotopic PL approximation of M from $\Psi(M)$, as noted in the main theorem, which follows and relies upon the notation for Ψ, as defined, above.

Theorem 6.1. *Let M be a compact, orientable,* C^2*-mainfold, with boundary. For any positive* $r < \delta$ *and for the previously defined mapping* $\Psi : M \to K(r)$, *the image*, $\Psi(M)$ *is ambient isotopic to M. Furthermore, there exists a PL ambient isotopic approximation of M and both of these sets can be made arbitrarily close to M.*

Proof. As a full proof would require extensive details, an outline of the critical steps follows.

The proof of the ambient isotopy between M and $\Psi(M)$ is similar to previous work by two of the present authors [5], but now has additional reliance upon Theorem 4.1 to ensure that $E_r(M)$ is $C^{1,1}$ and upon Corollary 4.2 to ensure appropriate ambient isotopic images of $E_r(M)$.

As already noted, the image $\Psi(M)$ need not be PL, since $\Psi(\partial M)$ need not be PL. However, since $\Psi(\partial M)$ is C^1, an ambient isotopic PL approximation $\alpha(\Psi(\partial M))$ of $\Psi(\partial M)$ can easily be constructed. Then, a PL approximation, $\beta(\Psi(M))$, to $\Psi(M)$ can be created by replacing $\Psi(\partial M)$ by $\alpha(\Psi(\partial M))$ and extending α to an ambient isotopy Ω between $\beta(\Psi(M))$ and $\Psi(M)$. The construction of Ω proceeds locally upon each triangle in $K(r)$ and on each line segment used to approximate subsets of $\Psi(\partial M)$. Then these local isotopies are pasted together via Lemma 5.2 to achieve the final desired ambient isotopic PL approximation to M, where this pasting is similar to other ambient isotopic approximation techniques in the literature [11,15,19]. □

7. Conclusion and challenges: Integration of theory and practice

This paper presents new theory for PL ambient isotopic approximation and reconstruction of C^2-surfaces with boundary. These theorems are expected to lead to improved algorithms, with preliminary algorithmic experiments already published [1].

The proofs rely upon the definition of the envelope of a manifold M. This envelope definition does not require M to be orientable, but the proof of an ambient isotopy PL approximation of M does utilize this orientability assumption. Experimental algorithms were created where this orientability requirement was dropped as an input criterion, with the results described, below. These experimental results may prompt even stronger theorems, as these images portray easily discernible improvements, as described, below.

In Fig. 3, there are four images. The top left is a surface reconstruction by merely feeding the sampled point cloud data of a Möbius strip directly to the power crust algorithm [3], which is not designed to directly accept such input. The top right image shows the approximation of the medial axis of the Möbius strip that is generated by the power crust algorithm. The lower left shows an envelope of the Möbius strip. The lower right shows the final reconstruction of this non-orientable surface as the internal medial axis of a reconstructed envelope. The improved image in the lower right was created by implementing a very short pre-processor to the existing power crust code and this technique warrants more exploration.

Acknowledgements

The authors sincerely thank the astute, anonymous referee for many insightful and constructive comments, which have led to significant improvements and to promising new directions for future work on non-orientable surfaces.

Appendix

The following terminology and notation is used throughout this paper.

Remark A.1. A function $f : M \to \mathbf{R}^3$ defined on a manifold M of dimension less than or equal to 2, is an (topological) embedding if $f : M \to f(M)$ is a homeomorphism (with respect to the subspace topology on $f(M)$). If, in addition, f is of class C^k on M ($k \geqslant 1$) and the Jacobian map of f is of full rank, then f is said to be a C^k embedding, and M (actually $f(M)$) is called an embedded C^k submanifold of $\mathbf{R}^3$.

In this article we present theoretical foundations for our work with computational models of curves and surfaces. We summarize the elements of differential geometry required to state and prove our results. Good treatments of this elementary material can be found in the texts [7,13].

We restrict our attention to curves and surfaces in three-dimensional Euclidean space. [8] Hereafter we assume that all differentiable objects are C^2, as defined below, unless otherwise stated (see [7]).

Definition A.1. A Hausdorff topological space M satisfying the second countability axiom is called a *C^2-differentiable manifold of dimension two* (*without boundary*) if it satisfies the following:

(1) For any point $x \in M$, there exists a pair (U, ϕ_U), where U is an open neighborhood of x in M, and $\phi_U : U \to A \subset \mathbf{R}^2$ is a homeomorphism of U with an open set of $\mathbf{R}^2$. The neighborhood U is called a coordinate neighborhood (or patch) of x and the function ϕ_U is called a coordinate function of x. The function ϕ_U introduces the *local coordinates* $\phi_U(x) = (u_1(x), u_2(x))$ for this patch. The pair (U, ϕ_U) is often referred to as a *coordinate patch*.
(2) For any coordinates patches $U,\ V$ with $U \cap V \neq \emptyset$, the map $\phi_V \circ (\phi_U)^{-1} : \phi_U(U \cap V) \to \phi_V(U \cap V)$ is C^2.

A C^2-differentiable manifold M of dimension two with boundary ∂M is defined similarly, as follows.

Definition A.2.

(1) If $x \in M - \partial M$, there is a coordinate pair as in (1) above. If $x \in \partial M$, there is a coordinate pair (U, ϕ_U) with a surjective homeomorphism $\phi_U : U \to H^2$, where H^2 is the half plane $\{(x_1, x_2) \in \mathbf{R}^2 :\ x_2 \geqslant 0\}$.
(2) Given two coordinates patches $U,\ V$ with $U \cap V \neq \emptyset$, the function $\phi_V \circ (\phi_U)^{-1} : \phi_U(U \cap V) \to \phi_V(U \cap V)$ is C^2 in the usual sense if $U \cap V$ contains no point in ∂M. Otherwise, the map $\phi_V \circ (\phi_U)^{-1}$ can be extended to a C^2-homeomorphism in an open subset of $\mathbf{R}^2$ that contains the domain $\phi_U(U \cap V)$.

[8] Generalizations can be found elsewhere [7].

If M is compact, ∂M is a disjoint union of finite closed curves, each of which is diffeomorphic [20] to the unit circle.

Let M be a two-dimensional manifold with or without boundary. A function $f : M \to \mathbf{R}^3$ is said to be a C^2-differentiable map if for any point $x \in M$, there is a coordinate patch (U, ϕ_U) about x so that the composition $f \circ (\phi_U)^{-1} : \phi(U) \to \mathbf{R}^3$ is C^2.

What we see as a surface in $\mathbf{R}^3$ in the conventional sense is the image of M in R^3 under f. In the case when M is a submanifold of $\mathbf{R}^3$, we often identify M with $f(M)$ if there is no risk of confusion. The map f is also called the parametrization of the surface. However, as is in the cases to follow, we often need to distinguish M and its image.

Since the Jacobian map $f_*(x)$ of f at $x \in M$ is of full rank 2, it gives rise to an injective linear map of the tangent space,[9] denoted TM_x, into the tangent space $T\mathbf{R}^3_{f(x)}$, which is identified with $\mathbf{R}^3$ in the conventional way.

The tangent space TM_x is identified with $\mathbf{R}^2$ with the standard coordinates (u_1, u_2) under the coordinate map ϕ_U. In terms of these coordinate systems, the matrix representation of $f_*(x)$ is the following three by two matrix:

$$\begin{pmatrix} \dfrac{\partial x_1}{\partial u_1} & \dfrac{\partial x_1}{\partial u_2} \\ \dfrac{\partial x_2}{\partial u_1} & \dfrac{\partial x_2}{\partial u_2} \\ \dfrac{\partial x_3}{\partial u_1} & \dfrac{\partial x_3}{\partial u_2} \end{pmatrix},$$

where $x_i(u_1, u_2) = f_i(u_1, u_2),\ i = 1, 2$, are the coordinate functions of f.

The image $f_*(x)(TM_x)$ is a plane passing through $f(x)$ in $\mathbf{R}^3$ and is called the *tangent plane* to the surface $f(M)$ at $f(x)$, but also referred to as the tangent plane to M at x. The ordinary dot product in $\mathbf{R}^3$ induces an inner product in the tangent plane. The induced inner product gives rise to the induced Riemannian metric in M. When we say a *surface* in $\mathbf{R}^3$, we implicitly understand the triple consisting of the manifold M, the embedding f and the induced Riemannian metric.

Let (M, f) be an embedded surface in $\mathbf{R}^3$. Denote by $\mathbf{n} = \mathbf{n}_x$ a (local) unit normal field along $f(M)$. Given a tangent vector X to M at x, $D_{f_*(X)}\mathbf{n}$ denotes the directional derivative of $\mathbf{n}$ in the direction of $f_*(X)$ in $\mathbf{R}^3$, where f_* is the Jacobian map of f at x. The derivative $D_{f_*(X)}\mathbf{n}$ is tangential to $f(M)$ at $f(x)$. By setting

$$D_{f_*(X)}\mathbf{n} = -f_*(AX),$$

one can obtain a linear operator A on the tangent space TM_x, see [13]. The map A determines the local geometric shape of the embedded surface $f(M)$ and A is a symmetric linear operator with respect to the induced Riemannian metric; hence, A can be represented by a 2×2 symmetric matrix with respect to any orthonormal basis for TM_x.

Definition A.3. The linear operator $A = A_x$ is called the *shape operator* (or the second fundamental form) of the surface (M, f). The eigenvalues of A are the *principal curvatures* of the surface at the point x (see, e.g., [13]).

Definition A.4. A point $x \in M$ is said to be a *critical* point of a C^2-function $g : M \to \mathbf{R}$ if the differential $\mathrm{d}g = (\partial g/\partial u_1)\,\mathrm{d}u_1 + (\partial g/\partial u_2)\,\mathrm{d}u_2 = 0$ at x, where (u_1, u_2) is a coordinate system about x in M. A critical point is called *non-degenerate* if its Hessian $Hg(x) = (\partial^2 g/\partial u_i \partial u_j)$ is invertible; otherwise it is called *degenerate*.

For our purposes, it is convenient to characterize the critical points of a function defined in M in the context of submanifolds, namely, in the extrinsic setting. Let g be as above. We state the following proposition without proof.

Proposition A.1. *The point $x \in M$ is a critical point of g if there is an open neighborhood U of $f(x)$ in $\mathbf{R}^3$ and a C^2-function $\tilde{g} : U \to \mathbf{R}$ with $\tilde{g} = g \circ f^{-1}$ such that the gradient $\nabla\tilde{g}$ in $\mathbf{R}^3$ is normal to the tangent plane to $f(M)$ at $f(x)$. Furthermore, such a (local) extension $\tilde{g}$ always exists.*

---

[9] The tangent space is an abstraction of the standard notion of a plane of tangent vectors for each point of a differentiable manifold in $\mathbf{R}^3$.

We now define the (global) energy function, G for a manifold with boundary by

$$G : M \times M \to \mathbf{R}, \quad G(x, y) = \|x - y\|^2, \tag{A.1}$$

where $\|x - y\|^2$ is the square of the ordinary distance function on $\mathbf{R}^6$.

We need to identify the critical points of G. In the intrinsic sense, a critical point is a pair $(x, y) \in M \times M$ such that $\mathrm{d}G(x, y) = 0$, as defined above. Extrinsically, recall that M is embedded by f into $\mathbf{R}^3$; hence, $M \times M$ is canonically embedded into $\mathbf{R}^6 = \mathbf{R}^3 \times \mathbf{R}^3$ under $f \times f : M \times M \to \mathbf{R}^3 \times \mathbf{R}^3$. Note also that the function G can naturally be extended in the entire $\mathbf{R}^3 \times \mathbf{R}^3$. Therefore, we may redefine, by Proposition A.1, a critical point of $G : M \times M \to \mathbf{R}$ to be a point $(x, y) \in M \times M$ where the gradient field $\nabla G(x, y)$ is normal to $(f \times f)(M \times M)$ at $(f \times f)(x, y)$.

Proposition A.2. *Let G be defined as in Eq.* (A.1). *Then, there exists a minimal positive critical value of G in* $M \times M$.

Proof. Obviously, $G(x, y) > 0$, for $x \neq y$. Second, note that G has a critical value $r > 0$, for example, the maximal value, since $M \times M$ is compact. The gradient of G in $\mathbf{R}^3 \times \mathbf{R}^3$ is given by

$$\nabla G = 2(x - y, -(x - y)),$$

where $x = (x_1, x_2, x_3),\ y = (y_1, y_2, y_3)$ are the standard Euclidean coordinates of $x,\ y$, respectively.

On the other hand, the tangent plane to $f(M)$ at $f(p),\ p \in M$ in $\mathbf{R}^3$ is spanned by two vectors $\partial f/\partial u_i,\ i = 1, 2$. Hence, the tangent space of $(f \times f)(M \times M)$ at $(x, y) = (f(p), f(q))$ in $\mathbf{R}^3 \times \mathbf{R}^3$ is the 4-space spanned by four vectors $(\partial f/\partial u_i)(p),\ i = 1, 2$ and $(\partial f/\partial v_i)(q),\ i = 1, 2$, where, as before, $(u_1, u_2),\ (v_1, v_2)$ denote local coordinates about $p,\ q$, respectively. The gradient ∇G is normal to the tangent space of $(f \times f)(M \times M)$ at $(f \times f)(p, q)$ if and only if

$$\sum_{k=1}^{3} (f_k(p) - f_k(q)) \frac{\partial f_k}{\partial u_i}(p) = 0, \quad i = 1, 2,$$

$$\sum_{k=1}^{3} -(f_k(p) - f_k(q)) \frac{\partial f_k}{\partial v_i}(q) = 0, \quad i = 1, 2.$$

If M has no boundary, this immediately tells us that (p, q) is a critical point of G if and only if either the line segment connecting $f(p), f(q)$ is normal to the tangent planes to $f(M)$ at $f(p)$ and $f(q)$ in $\mathbf{R}^3$, or $f(p) = f(q)$. We claim that if

$$c = \inf\{r > 0 : r \text{ is a critical value of } G\} \tag{A.2}$$

then c is positive, with elementary proofs available [22,23].

When M has a non-empty boundary, the situation is slightly more complicated. There will be three possible cases for critical points to occur. (1) (x, y) is a critical point of G and x and y both lie in the interior of M; (2) (x, y) is a critical point G and one of them lies in the interior of M and the other lies in ∂M; (3) (x, y) is a critical point of G and x and y both lie in ∂M. In any of these cases, a slightly modified proof for the case without boundary also works. □

References

[1] K. Abe, J. Bisceglio, T.J. Peters, A.C. Russell, D.R. Ferguson, T. Sakkalis, Computational topology for reconstruction of surfaces with boundary: integrating experiments and theory, in: Proc. IEEE Int. Conf. on Shape Modeling and Applications, June 15–17, 2005, Cambridge, MA, IEEE Computer Society, Los Alamitos, CA, pp. 288–297.

[2] N. Amenta, S. Choi, T. Dey, N. Leekha, A simple algorithm for homeomorphic surface reconstruction, in: ACM Symp. on Computational Geometry, 2000, pp. 213–222.

[3] N. Amenta, S. Choi, R. Kolluri, The power crust, in: Proc. Sixth ACM Symp. on Solid Modeling, 2001, pp. 249–260.

[4] N. Amenta, S. Choi, R. Kolluri, The power crust, union of balls and the medial axis transform, Comput. Geom. Theory Appl. 19 (2001) 127–173.

[5] N. Amenta, T.J. Peters, A.C. Russell, Computational topology: ambient isotopic approximation of 2-manifolds, Theoret. Comput. Sci. 305 (2003) 3–15.

[7] W.M. Boothby, An Introduction to Differentiable Manifolds and Riemannian Geometry, second ed., Academic Press, New York, 1986.

[9] T.K. Dey, S. Goswami, Tight Cocone: a water-tight surface reconstructor, in: Proc. Eighth ACM Symp. on Solid Modeling and Applications, 2003, pp. 127–134.
[10] M.P. doCarmo, Differential Geometry of Curves and Surfaces, Prentice-Hall, Englewood Cliffs, NJ, 1976.
[11] D. Freedman, Combinatorial curve reconstruction in Hilbert spaces: a new sampling theory and an old result revisited, Comput. Geom. Theory Appl. 23 (2) (2002) 227–241.
[12] M. Gopi, On sampling and reconstructing surfaces with boundaries, in: S. Wismath (Ed.), Proc. Canadian Conf. on Computational Geometry, 2002, pp. 49–53.
[13] N.J. Hicks, Notes on Differential Geometry, Van Nostrand Math. Studies #3, Princeton, NJ, 1965.
[14] M.W. Hirsch, Differential Topology, Springer, New York, 1976.
[15] T. Maekawa, N.M. Patrikalakis, T. Sakkalis, G. Yu, Analysis and applications of pipe surfaces, Comput. Aided Geom. Design 15 (5) (1998) 437–458.
[16] J. Milnor, Morse Theory, Princeton University Press, Princeton, NJ, 1969.
[17] G. Monge, Application de l'Analys à la Geometrie, Bachelier, Paris, 1850.
[18] E.L.F. Moore, Computational Topology of Spline Curves for Geometric and Molecular Approximations, Doctoral Dissertation, The University of Connecticut, 2006.
[19] E.L.F. Moore, T.J. Peters, Computational topology for geometric design and molecular design, in: D.R. Ferguson, T.J. Peters (Eds.), Mathematics for Industry: Challenges and Frontiers, SIAM, Philadelphia, 2005, pp. 125–137.
[20] B. O'Neill, Elementary Differentiable Geometry, Academic Press, New York, 1972.
[21] W. Rudin, Principles of Mathematical Analysis, second ed., McGraw-Hill, New York, 1961.
[22] T. Sakkalis, T.J. Peters, Ambient isotopic approximations for surface reconstruction and interval solids, ACM Symp. on Solid Modeling, Seattle, June 9–13, 2003.
[23] T. Sakkalis, T.J. Peters, J. Bisceglio, Isotopic approximations and interval solids, CAD 36 (11) (2004) 1089–1100.
[24] S. Willard, General Topology, Addison-Wesley Publishing Company, Reading, MA, 1970.
[25] F.-E. Wolter, Cut locus & medial axis in global shape interrogation & representation, MIT Design Laboratory Memorandum 92-2, December 1993 (revised version).