# The Persistent Morse Complex Segmentation of a 3-Manifold [*]

Herbert Edelsbrunner[1,2,3] and John Harer[4,5]

[1] IST Austria (Institute of Science and Technology Austria).
[2] Dept. Comput. Sci., Duke Univ., Durham, North Carolina, USA.
[3] Geomagic, Research Triangle Park, North Carolina, USA.
[4] Dept. Math., Duke Univ., Durham, North Carolina, USA.
[5] Program Comput. Bio. Bioinf., Duke Univ., Durham, North Carolina, USA.

**Abstract.** We describe an algorithm for segmenting three-dimensional medical imaging data modeled as a continuous function on a 3-manifold. It is related to watershed algorithms developed in image processing but is closer to its mathematical roots, which are Morse theory and homological algebra. It allows for the implicit treatment of an underlying mesh, thus combining the structural integrity of its mathematical foundations with the computational efficiency of image processing.

## 1 Introduction

The extraction of shape information from density functions is an important topic in medical imaging. We formalize this problem as the construction and simplification of the Morse complex of the function. Before discussing this formalization and our algorithm, we discuss our motivation and briefly survey the extensive related literature.

**Motivation and prior work.** Medicine uses a variety of technologies to non-invasively obtain images of the anatomy of a subject. Examples are magnetic resonance imaging, X-ray and other types of computed tomography, ultrasound, and X-ray projectional radiography. They provide information about the anatomy expressed in terms of intensity or density. To make sense of the data, we need image segmentation algorithms that extract shapes by delineating their boundaries. This is a necessary first step for a multitude of medical tasks, including the quantification of tissue, the diagnosis, and the study of anatomic structure.

The segmentation of a three-dimensional image can be approached in a variety of ways. Today, many different algorithms are in use, each with its own strengths and weaknesses. We refer to Clarke *et al.* [4] and Pham, Xu and Prince [19] for surveys of segmentation algorithms in medical image analysis. In an attempt to organize the body of prior work, we distinguish between *direct* and *indirect* approaches to segmentation. In the direct approach, an algorithm generates shapes using local considerations based on the density function. An example is *thresholding*, in which a small number of real

thresholds is used to partition space into the preimages of intervals between contiguous thresholds [21]. Another example is *region growing*, in which seed points are grown into regions using local expansion criteria [10]. A third example works by *spitting and merging* regions [12]; see [14] for an extension to three-dimensional magnetic resonance images. The recently most successful of the indirect approaches are *deformable models* [16] or *level set methods* [23], in which boundary surfaces are computed through numerical solutions to differential equations. Shapes can also be 'learned' using *artificial neural networks* [11] or other adaptive structures developed within artificial intelligence. Sometimes, an *atlas* representing an idealized or expected segmentation is available, and the problem reduces to *registering* the atlas with the image [13].

The prior work most similar to ours is on the *watershed method*, which is another direct approach to segmentation. It is easiest explained by the analogous process of flooding a landscape in which water seeps in from every local minimum. As the water rises, it fills up a basin around each minimum and when two basins meet, it erects a dam that follows the line along which the water meets. Algorithms in image processing that construct the dams defining the segmentation can be found in [3, 25]. The concept itself is however older within image processing [2, 9]. A survey of watershed algorithms can be found in [20]. The algorithm of Vincent and Soilles [25] has been extended to three-dimensional magnetic resonance images by Sijbers *et al.* [22]. The authors of the latter paper cope with over-segmentation by applying a diffusion filter [26] eliminating many spurious local minima before running the watershed algorithm. They also merge regions in a postprocessing phase, using a statistical test to guide the process.

**Results.** The main contribution of this paper is a new segmentation algorithm. It takes the direct approach and falls into the class of watershed algorithms, constructing regions by following the gradient of the density function. The algorithm has a short list of distinguishing characteristics based on old and new work in differential and algebraic topology:

1. The basic segmentation of space is achieved by constructing what we call the Morse complex of the density function [7, 24].
2. We coarsen the complex without altering its mathematical properties using homological persistence to guide the process [8].
3. We find a compromise between structural integrity, numerical accuracy, and computational efficiency in which none of the three measures makes substantial sacrifices.

The two-dimensional version of our algorithm has been implemented and used to decompose 2-manifolds describing mechanical shapes [6]. The success of that algorithm, even for rather complicated shapes, motivates us to extend the algorithm to three dimensions. Similar to its two-dimensional counterpart, our algorithm works for general 3-manifolds and not just for $\mathbb{R}^3$. It is most efficient when the data permits an implicit representation of the triangulation representing the 3-manifold.

**Outline.** Section 2 presents background from differential topology and from combinatorial topology. Section 3 explains the segmentation algorithm. Section 4 discusses the simplification of the segmentation. Section 5 presents plans for the future.

## 2 Background

In this section, we review the mathematical background needed for our algorithm. We discuss Morse functions, piecewise linear maps, and persistent homology.

**Morse functions.** Letting $f : \mathbb{M} \to \mathbb{R}$ be a smooth function on a 3-manifold, we get the derivative at a point $x \in \mathbb{M}$ as a linear map from the tangent space to the reals, $\mathrm{d}f_x : \mathrm{TM}_x \to \mathbb{R}$. The point $x$ is *critical* if $\mathrm{d}f_x$ is the zero map, otherwise, it is *regular*. A *critical value* is the image $f(x)$ of a critical point $x$. Assuming local coordinates at $x$, the *Hessian* is the matrix $H(x)$ of second-order partial derivatives. It distinguishes *non-degenerate* critical points, for which the Hessian is invertible, from *degenerate* ones, for which $\det H(x) = 0$. In this paper, we are interested in using critical point theory for smooth functions merely as a guiding intuition, so we focus on the generic case of *Morse functions* defined by the following two requirements:

I. All critical points are non-degenerate.
II. All critical points have unique function values.

The Morse Lemma asserts that a non-degenerate critical point $x$ permits local coordinates such that $f(x_1, x_2, x_3) = f(x) \pm x_1^2 \pm x_2^2 \pm x_3^2$ in a neighborhood of $x$ [15, 17]. This implies that non-degenerate critical points are isolated and, if $\mathbb{M}$ is compact, any Morse function on $\mathbb{M}$ has only finitely many critical points. The number of minus signs in the above expression is independent of the chosen local coordinates and equals the number of negative eigenvalues of the Hessian. It is referred to as the *index* of $x$ and distinguishes *minima* (index 0), 1-*saddles* (index 1), 2-*saddles* (index 2), and *maxima* (index 3) from each other. We get a local picture by drawing a small sphere around a critical (or regular) point $x$, and intersecting it with the level set to separate the region below $x$ (the set of points $y$ with $f(y) < f(x)$) from the region above $x$ (points $y$ with $f(y) > f(x)$). Non-degenerate critical points have the characteristic local pictures shown in Figure 1.



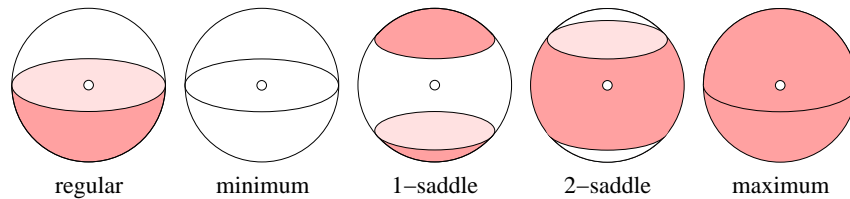| regular | minimum | 1−saddle | 2−saddle | maximum |

Fig. 1: Local pictures of regular and critical points of a Morse function on a 3-manifold. The level set separates the shaded region below $x$ from the white region above $x$.

Critical points are often used to study the topology of the manifold, by sweeping $\mathbb{M}$ in the direction of increasing function value and monitoring the *sub-level set* defined as $\mathbb{M}_a = \{x \in \mathbb{M} \mid f(x) \leq a\}$, where $a \in \mathbb{R}$ is the value of the current level.

As $a$ increases, we observe topology changes whenever $a$ passes a critical value. The changes depend on the indices of the corresponding critical points and accumulate to give $\chi = \#\mathrm{min} - \#\mathrm{sad}_1 + \#\mathrm{sad}_2 - \#\mathrm{max} = 0$ for the *Euler characteristic* of $\mathbb{M}$.

**Morse complexes.** Assuming a Riemannian metric on $\mathbb{M}$, we can use local coordinates to define the *gradient* as the vector $\nabla f(x)$ of first-order partial derivatives at $x$. If we follow the flow defined by the gradient, we trace out an *integral curve* $\gamma : \mathbb{R} \to \mathbb{M}$ that begins at a critical point, the *origin*, $\mathrm{org}\, \gamma = \lim_{t \to -\infty} \gamma(t)$, and ends at another critical point, the *destination*, $\mathrm{dest}\, \gamma = \lim_{t \to \infty} \gamma(t)$, without containing either. Two regular points either belong to the same or to two disjoint integral lines, which implies that the integral lines together with the critical points partition the manifold. We can now form a coarser partition by merging all integral lines with the same destination. Specifically, the *descending manifold* of a critical point $x$ is

$$D(x) = \{x\} \cup \{y \in \mathrm{im}\, \gamma \mid \mathrm{dest}\, \gamma = x\}.$$

A minimum cannot be the destination of any integral line; its descending manifold is therefore just a point, the minimum itself. The descending manifold of a 1-saddle consists of two integral lines approaching the 1-saddle from opposite sides in the direction of the eigenvector with negative eigenvalue. Together with the 1-saddle, these two integral lines form an open 1-manifold whose endpoints are the origins of the two curves, which may or may not be the same. Similarly, the descending manifold of a 2-saddle is an open disk and that of a maximum is an open ball. Symmetrically, we define the *ascending manifold* of a critical point as the point $x$ itself union all integral lines with origin at $x$. A Morse function is *Morse-Smale* if

   III. the descending and ascending manifolds intersect transversally.

Being Morse-Smale is a generic property of smooth functions. Property III implies that the descending manifolds are the *cells* of a complex, in the sense that the cells partition $\mathbb{M}$ and the boundary of each cell is a union of lower-dimensional cells in the complex. We call this the *Morse complex* of the function.

**Homology.** The *$p$-th homology group* is an algebraic representation of the collection of $p$-dimensional holes of a topological space. Formally, it is the group of $p$-dimensional cycles divided by the group of $p$-dimensional boundaries, $\mathsf{H}_p = \mathsf{Z}_p / \mathsf{B}_p$ [18]. The *$p$-th Betti number* is the rank of this group, $\beta_p = \mathrm{rank}\, \mathsf{H}_p$, and can be interpreted as the number of $p$-dimensional holes. The only possibly non-zero Betti numbers a 3-manifold can have are for the dimensions $p = 0, 1, 2, 3$. The Euler-Poincaré formula asserts that

$$\chi = \beta_0 - \beta_1 + \beta_2 - \beta_3, \tag{1}$$

which should be compared with (3). As suggested by the two equations, we can make index-$p$ critical points responsible for the $p$-dimensional holes counted by the $p$-th Betti number. It is convenient to use reduced homology groups [18, Chapter 1] to rationalize the classification of critical vertices. The ranks of these groups are the same as their non-reduced counterparts, except for $\tilde{\beta}_0 = \beta_0 - 1$ if the space is non-empty and $\tilde{\beta}_{-1} = 1$

| type | index | $\tilde{\beta}_{-1}$ | $\tilde{\beta}_0$ | $\tilde{\beta}_1$ | $\tilde{\beta}_2$ |
|------|-------|------|------|------|------|
| regular | | 0 | 0 | 0 | 0 |
| minimum | 0 | 1 | 0 | 0 | 0 |
| 1-saddle | 1 | 0 | 1 | 0 | 0 |
| 2-saddle | 2 | 0 | 0 | 1 | 0 |
| maximum | 3 | 0 | 0 | 0 | 1 |

Table 1: For simple vertex types, the dimension of the non-zero reduced Betti number of the lower link, if any, is one less than the index of the vertex.

if the space is empty. Table 1 gives the relation between the type of a vertex and the reduced Betti number of its lower link.

**Piecewise linear manifolds.** From now on, we consider the case in which the 3-manifold, $\mathbb{M}$, is the underlying space of a simplicial complex $K$ and that $f : \mathbb{M} \to \mathbb{R}$ is obtained by piecewise linear extension of function values specified at the vertices of $K$. Thinking of $f$ as the approximation of a smooth function, we introduce corresponding concepts. First we need some definitions. The *star* of a vertex $u$ is the set of simplices that contain the vertex, $\mathrm{St}\, u = \{\sigma \in K \mid u \leq \sigma\}$. The *link* of $u$ is the set of faces of simplices in the star that do not contain $u$, $\mathrm{Lk}\, u = \{\tau \leq \sigma \in \mathrm{St}\, u \mid \tau \notin \mathrm{St}\, u\}$. Since $K$ triangulates a 3-manifold, every vertex star has the topology of an open ball and every vertex link has the topology of a sphere. The *lower star* and *lower link* are the subcomplexes induced by the vertices below $u$:

$$\mathrm{St}_- u = \{\sigma \in \mathrm{St}\, u \mid u \neq v \leq \sigma \Rightarrow f(v) < f(u)\};$$
$$\mathrm{Lk}_- u = \{\tau \in \mathrm{Lk}\, u \mid v \leq \tau \Rightarrow f(v) < f(u)\}.$$

Assuming the vertices in $K$ have pairwise different function values, the level set that passes through $u$ intersects the link in a (not necessarily connected) curve that avoids all vertices and decomposes the link into two (not necessarily connected) regions, one below and the other above $u$. The region below $u$ contains the lower link and has the same homotopy type, which can be proved by establishing a deformation retraction from the region to the lower link. Guided by the smooth case, we call $u$ a *regular vertex*, a *minimum*, a *1-saddle*, a *2-saddle*, and a *maximum* if its lower link has the homotopy type of the corresponding shaded region in Figure 1. A vertex that does not fit this classification has $\tilde{\beta}_0 + \tilde{\beta}_1 \geq 2$, with $\tilde{\beta}_0 + 1$ the number of components in the lower link and $\tilde{\beta}_1 + 1$ the number of components in the symmetrically defined upper link. We thus think of it as the simultaneous embodiment of a $\tilde{\beta}_0$-fold 1-saddle and a $\tilde{\beta}_1$-fold 2-saddle.

Our assumption of vertices with pairwise different function values implies that every simplex of $K$ belongs to a unique lower star. In other words, the lower stars partition $K$. Indexing the vertices in the order of increasing function value, $f(u_1) < f(u_2) < \ldots < f(u_m)$, we may form a nested sequence of subcomplexes,

$$\emptyset = K_0 \subset K_1 \subset K_2 \subset \ldots \subset K_m = K, \tag{2}$$

called a *filtration* of $K$, where $K_j = \bigcup_{i=1}^{j} \mathrm{St}_- u_i$. The complex $K_j$ has the same homotopy type as the sub-level set $\mathbb{M}_a$ for every $f(u_j) \leq a < f(u_{j+1})$, which can be proved by establishing a deformation retraction from $\mathbb{M}_a$ to $K_j$. The difference between two contiguous complexes is the extra lower star, $K_{j+1} - K_j = \mathrm{St}_- u_{j+1}$. By definition, the Euler characteristic of $K_j$ is the alternating sum of simplex numbers. We claim it is also the alternating sum of critical point numbers,

$$\begin{aligned}
\chi &= \#\mathrm{vert} - \#\mathrm{edg} + \#\mathrm{tri} - \#\mathrm{tet} \\
&= \#\mathrm{min} - \#\mathrm{sad}_1 + \#\mathrm{sad}_2 - \#\mathrm{max},
\end{aligned} \tag{3}$$

where a multiple critical point is counted $\tilde{\beta}_1 - \tilde{\beta}_0$ times. Indeed, the equation holds initially, for $j = 0$, and it is maintained when we add the lower star of $u_{j+1}$ to $K_j$.

**Persistence.** We use homology groups to count the holes in the complexes $K_i$ ordered by inclusion as in (2). Fixing the dimension to $p$, we write $F_i = \mathsf{H}_p(K_i)$ for the $p$-th homology group of the $i$-th complex. The chain of inclusions among the complexes implies a chain of maps among the homology groups induced by inclusion,

$$F_0 \to F_1 \to F_2 \to \ldots \to F_m.$$

Let $f_i^j : F_i \to F_j$ be the composition of $j - i$ of the maps along the chain. The corresponding *persistent homology group* is the image of this map, $F_i^j = \mathrm{im}\, f_i^j$, and the corresponding *persistent Betti number* is $\beta_i^j = \mathrm{rank}\, F_i^j$. We use these numbers to define the life-time of individual homology classes. The description of this idea is made complicated by multiple critical points, which may simultaneously create several classes and destroy several other classes. Defining

$$\mu_i^j = (\beta_i^{j-1} - \beta_i^j) - (\beta_{i-1}^{j-1} - \beta_{i-1}^j),$$

we say that $\mu_i^j$ homology classes *live* from $f(u_i)$ to $f(u_j)$, *being born* at time $f(u_i)$ and *dying* at time $f(u_j)$. Indeed, the first difference can be interpreted as the number of classes born at or before time $f(u_i)$ that die after time $f(u_{j-1})$ but at or before time $f(u_j)$. With a similar interpretation of the second difference, $\mu_i^j$ is the number of classes born some time in $(f(u_{i-1}), f(u_i)]$ that die some time in $(f(u_{j-1}), f(u_j)]$. Since there is no activity in the open interval between contiguous critical values, we can attribute the births and deaths as stated. The *persistence* of a class counted by $\mu_i^j$ is $f(u_j) - f(u_i)$. An algorithm that pairs up births with deaths and computes the persistence of homology classes can be found in [8]. Using the integers modulo 2 as the coefficient group for homology, it takes time $\mathrm{O}(N^3)$ in the worst case for a filtration of a complex with $N$ simplices. The experimentally observed running time is much less than cubic and seems to be close to linear in $N$. The persistence of the maxima and the minima can be computed faster, in worst case time close to linear in $N$.

## 3 Algorithm

In this section, we present the segmentation algorithm. After explaining its global structure, we describe its actions in the simple cases. The general case can be interpreted as an accumulation of simple cases, as given in the appendix.

**Problem specification.** As input to the algorithm, we assume a triangulation $K$ of the 3-manifold $\mathbb{M}$ and a function value for each vertex. Writing $u_1$ to $u_m$ for the vertices, we assume pairwise different function values, $f(u_i) \neq f(u_j)$ whenever $i \neq j$, and we justify the assumption by the use of a simulated perturbation [5, Chapter 1.4]. As explained earlier, the function values can be extended linearly over all simplices to give a continuous PL function $f : |K| \to \mathbb{R}$. The output is a marking of some simplices and a labeling of all unmarked simplices. Specifically, the *marking* is a map $\mu : K \to \{0, 1\}$ and the subset of *unmarked* simplices is $K^\circ = \mu^{-1}(0)$. The *labeling* is a map $\lambda : K^\circ \to [1, m]$ whose image is the set of subscripts of the maxima. The interpretation of $\mu$ and $\lambda$ is as follows. The subset of *marked* simplices, $\bar{K} = \mu^{-1}(1)$, corresponds to the union of descending 0-, 1-, and 2-manifolds in the smooth category. It decomposes the rest into components $K^\ell = \lambda^{-1}(\ell)$, which correspond to the (3-dimensional) cells of the Morse complex. The simplices of each component are labeled with the subscript of the generating maximum.

**Sweep from the top.** Before running the algorithm proper, we sort the vertices of the triangulation and re-name them such that $f(u_1) < f(u_2) < \ldots < f(u_m)$. The algorithm sweeps the triangulation from top to bottom, in the order of decreasing function value. Recall that $K_i$ is the union of lower stars of the vertices $u_1$ to $u_i$. Because of the chosen ordering, the algorithm explores the complement, $L_i = K - K_i$, before it explores $K_i$. We call a vertex in $K_i$ *stained* if it belongs to a marked simplex in $L_i$. The action taken at a vertex depends on whether it is regular or critical.

```
        for i = m downto 1 do
            case u_i is a maximum:
                label all simplices in the star of u_i with i
            case u_i is regular:
1               construct a spanning tree of the stained vertices in the lower link of u_i;
2               mark the simplices in the lower star incident to the spanning tree;
3               label the other simplices in the lower star by copying from neighbors
            case u_i is a 1-saddle:
                do Steps 1 to 3 for each component of the lower link
            case u_i is a 2-saddle:
                construct a spanning cactus whose cycle separates the poles;
                do Steps 2 and 3 substituting the cactus for the tree
            case u_i a minimum:
                mark u_i
        endfor.
```

A *cactus* is a tree plus a single extra edge, which thus defines a unique cycle. To construct a cactus as needed in the 2-saddle case, we first construct a spanning tree of all vertices in the lower link, we add an edge to form a cycle, and we prune the cactus while retaining the cycle and the stained vertices.

**Invariants.** Before discussing the cases in more detail, we make general observations and formulate properties that hold throughout the algorithm. Note first that a simplex is marked or labeled when it is first encountered, which is when the algorithm processes

its highest vertex. Since $K_i$ consists of all simplices that have no vertex with subscript higher than $i$, the set of marked or labeled simplices right before processing $u_i$ is its complement, $L_i = K - K_i$. While $L_i$ is generally not closed, it is *relatively closed* in the sense that every face of a simplex in $L_i$ is either in $L_i$ or in $K_i$. This is of course trivially true as $L_i$ and $K_i$ partition the entire complex. A more interesting observation is that the set of marked simplices shares the same property, that is, every face of a marked simplex in $L_i$ is either also marked or belongs to $K_i$. We now state this and two other claims as invariants of the algorithm.

INVARIANTS. The following relations hold for all subscripts of vertices, $i$, and all subscripts of maxima, $\ell$.
   (i) The set $\bar{L}_i = L_i \cap \bar{K}$ is relatively closed.
   (ii) The set $L_i^\circ = L_i - \bar{L}_i = \bigcup_\ell L_i^\ell$ is open.
   (iii) The set $L_i^\ell = L_i \cap K^\ell$ is empty or an open ball.

Since the algorithm works without back-tracking, $\bar{L}_i$ is the set of simplices marked right before processing $u_i$. To explain (ii), we call a set *open* if the inclusion of a simplex implies the inclusion of its entire star. It is then not difficult to see that (ii) holds. With a little bit more effort, we can also see that each label defines an open ball, as claimed in (iii). To see what this means, we note that $L_i^\ell$ is the subset of already processed simplices in the set $K^\ell = \lambda^{-1}(\ell)$, which represents a cell in the Morse complex. We call $L_i^\ell$ an *open ball* if the union of interiors of its simplices is homeomorphic to $\mathbb{R}^3$. To know that the cells are open balls is not important to understand the algorithm, but it is important to understand the result of the algorithm. For example, it implies that a marked vertex can be isolated only if the entire Morse complex consists of a single cell and the vertex forms the boundary of that cell. In this case, $K$ triangulates the 3-sphere and the function $f$ has only two critical vertices, a minimum and a maximum.

**Regular case.** Recall that $u = u_i$ is regular iff the level set defined by $u$, $f^{-1}(f(u))$, intersects the link in a single closed curve decomposing the link into two open disks. The lower link is a deformation retract of one of these disks and is therefore contractible. When we process $u$, we mark and label the simplices in its lower star. It is convenient to represent each such simplex by its face in the lower link, which has one lower dimension; see Figure 2. We now describe the three steps taken by the algorithm.

Step 1. Construct a spanning tree within the lower link that touches all its vertices. Extract a subtree $T_u$ that touches all stained vertices.

Step 2. Mark the triangles and edges in the lower star whose faces in the lower link are the edges and vertices of $T_u$.

The tree $T_u$ decomposes the lower link into regions we call *bays*, each adjacent to the complement of the lower link along edges we call its *frontier*. A triangle on the other side of the frontier corresponds to a tetrahedron in the star that does not belong to the lower star and is therefore already labeled. The labels we find across the frontier of a bay are all the same.

Step 3. Traverse each bay of the lower link, from the frontier inwards, and copy the label found outside the frontier to the corresponding tetrahedra, triangles, and edges in the lower star.
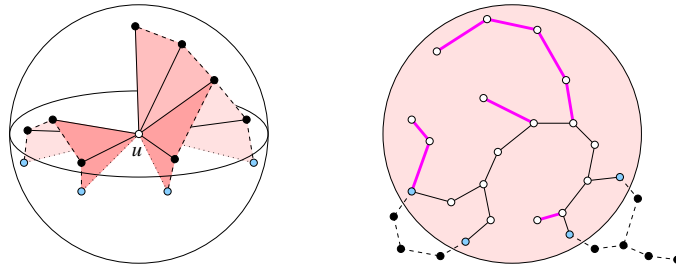
Fig. 2: Left: the star and link of the vertex $u = u_i$. The (solid) marked vertices lie in the upper hemi-sphere and the (shaded) stained vertices lie in the lower hemi-sphere. Right: the lower link of $u$ drawn as a disk in the plane. The (solid black) spanning tree of the stained vertices is a subtree of the (solid black and shaded) spanning tree of the lower link.

An additional layer of detail about the three steps will be given in our description of the general case in the appendix, which subsumes the regular point as a special case.

**2-saddle case.** Recall that the vertex $u = u_i$ is a 2-saddle iff $f^{-1}(f(u))$ intersects the link in two closed curves decomposing it into two open disks and an open annulus, as illustrated in Figure 3, where the lower link is a deformation retract of the annulus. The
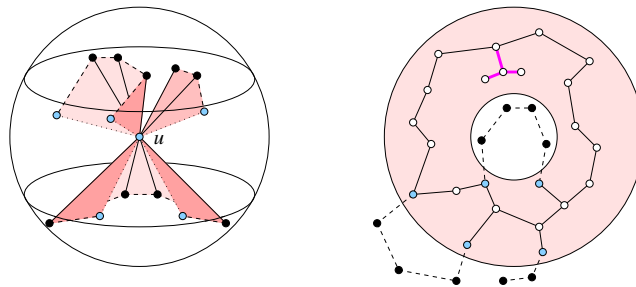


Fig. 3: Left: the star and link of the vertex $u = u_i$. The (solid) marked vertices lie on the polar caps and the (shaded) stained vertices lie on the equatorial belt that separates the caps. Right: the lower link of $u$ drawn as an annulus in the plane. The (solid black) spanning cactus of the stained vertices is a sub-cactus of the (solid black and shaded) spanning cactus of the lower link.

three steps processing $u$ are similar to those in the regular case, except that we begin with a cactus rather than a tree. We describe the first step in which this difference is most important.

Step 1. Construct a spanning tree within the lower link that touches all its vertices. Add a single edge creating a cycle that goes around the annulus, separating the two disks. Extract a cactus $T_u^+$ that touches all stained vertices.

By virtue of being a cactus, $T_u^+$ contains the cycle created by the edge added to the spanning tree. Steps 2 and 3 are the same as in the regular case.

**Implementation and running time.** The segmentation algorithm allows for a variety of short-cuts that make the implementation both easier and more efficient. The biggest gain can be expected from avoiding any explicit construction of the three-dimensional triangulation, the complex $K$. This is possible in the common case in which the data points form a regular grid in $\mathbb{R}^3$. However, the short-cuts do not affect the asymptotic running time of the algorithm, which we now address.

Let $m$, $n$ and $t$ be the number of vertices, edges, and triangles in $K$ and write $n_i$ and $t_i$ for the number of edges and triangles in the link of the vertex $u_i$. The asymptotic running time of the algorithm depends on the implementation of $\texttt{Step 1}$ and, in particular, on the construction of the spanning trees and spanning cacti. An attractive way to resolve ambiguities prefers short edges over long edges, thus constructing minimum spanning trees and cacti by always adding the shortest edge available. Every step of the construction for $u_i$ takes time at most some constant times $n_i \log_2 n_i$, the time it takes to sort the edges by length. To take the sum over all vertices, we observe that the total number of triangles is $t = \frac{1}{2} \sum_i t_i$, because each triangle belong to two vertex links. Since every link is a triangulated sphere, we have $n_i = \frac{3}{2} t_i$ for each $i$ and therefore $\sum_i n_i = 3t$. Furthermore, $t \leq n$ because every triangle has three edges and every edge belongs to at least three triangles. The total running time is therefore some constant times $\sum_i n_i \log_2 n_i \leq 3n \log_2(\max_i n_i)$. For the common case in which each vertex has only a constant size star, the algorithm thus runs in time proportional to the data size.

## 4 Coarsening

Over-segmentation is a consequence of the algorithm creating a cell for each and every local maximum. In this section, we discuss a greedy approach to progressively coarsening the initial, fine segmentation until a desired level of resolution is reached.

**Turning off maxima.** A coarsening of the segmentation is achieved by turning off a subset of the maxima and treating them as regular points. The cell of an off maximum is merged into another cell, which is accomplished by turning off a 2-saddle whose descending 2-manifold would otherwise separate the cells. As a first approximation, we can understand the result by running the segmentation algorithm with some modifications in how it treats 2-saddles. Assuming an arbitrary but fixed assignment of the maxima to on and off, the algorithm takes different actions at a 2-saddle, $u$, depending on the status of the cells that meet at $u$. Let $i$ and $j$ be the labels of their tetrahedra (the subscripts of the maxima that generate the two cells).

$\texttt{Case A.}$ If $i = j$ or $i \neq j$ and $u_i$ and $u_j$ are both on, then we treat $u$ the same way as in the original algorithm, starting a locally separating descending 2-manifold at $u$.

$\texttt{Case B.}$ If $i \neq j$ and $u_j$ is off, then we merge the cell of $u_j$ into that of $u_i$ by changing the label of its simplices from $j$ to $i$ and suppressing the creation of a separating descending 2-manifold at $u$.

In summary, the algorithm determines the status of a 2-saddle (on in `Case A` and off in `Case B`) based on the status of the maxima generating the cells that meet at the 2-saddle. A drawback of this algorithm is that it fails to make the same distinction for 1-saddles and minima, and it is indeed impossible to make the appropriate assignment based on the local information available when we sweep the function from top to bottom.

**Collapsing.** We remedy this deficiency by making the assignment in an order that is independent of the sweep direction. To describe this, we observe that whenever we turn off a maximum, we also turn off a 2-saddle, effectively cancelling the two critical points. We implement this operation by a *collapse in the large*, which unmarks the descending 2-manifold of the 2-saddle and re-labels the cell of the maximum. A collapse in the large translates into a sequence of collapses in the small, each affecting a pair of simplices in the triangulation. For example, to remove the cell, we start the process by initializing an open ball to a tetrahedron in the star of the maximum. The ball is then expanded by collapsing its complement until the cell is exhausted. For the 2-manifold, we start the process by initializing an open disk to a marked triangle in the star of $u$ that does not belong to any other descending 2-manifold. By construction, the triangle connecting $u$ to the bridge completing the cycle in its link is such a triangle. This disk is expanded by collapsing its marked complement. The process either exhausts the entire descending 2-manifold or comes to a halt where the 2-manifold merges with other descending 2-manifolds.

In the process of collapsing 2-manifold/cell pairs, we may deteriorate the surrounding of a descending 1-manifold until it belongs to only one descending 2-manifold. At this moment, we collapse the 1-/2-manifold pair. Similarly, when a descending 0-manifold (a minimum) finds itself part of only one descending 1-manifold, we collapse the 0-/1-manifold pair. As before, each collapse in the large translates into a sequence of collapses in the small. The process continues until no further collapses can be applied. The result of this algorithm may depend on the sequence of collapses, for example if a 2-manifold can be collapsed from two different 1-manifolds in its boundary. We get a unique result by prioritizing the collapses. Most naturally, we would use the function value of the lower-index critical point of the defining pair as priority and perform collapses with higher priority before collapses with lower priority.

**Greedily following persistence.** The above procedure defines $2^{\#\max} - 1$ different segmentations, one for each non-empty subset of the maxima. To rationalize the choice, we suggest to turn off maxima greedily, following the persistence measure as introduced in [8]. This is the time-lag between when a cell is initialized and when it first meets another, older cell. We consider the union of these two cells as the continuation of the older cell. The measure of a maximum, $v$, is therefore $\pi(v) = f(v) - f(w)$, where $w$ is the highest 2-saddle that separates the cell of $v$ from the cell of a maximum higher than $v$. This measure can be computed in time proportional to $n\alpha(n)$ using a union-find data structure that keeps track of the components of $L_i$ during a top to bottom sweep. Given the initial segmentation (with all maxima turned on) and the measure $\pi$ at every maximum, the greedy algorithm turns off maxima in the sequence of increasing measure. After re-naming the maxima such that $\pi(v_1) < \pi(v_2) < \ldots < \pi(v_{\#\max})$ and letting

$w_1, w_2, \ldots, w_{\#\max}$ be the corresponding sequence of 2-saddles, we can now state the greedy algorithm more formally.

```
for i = 1 to #max − 1 do
    collapse 2-manifold/cell pair of wᵢ and vᵢ;
    repeat
            collapse 1-/2- and 0-/1-manifold pairs
    until no further collapses are possible
endfor.
```

Of course, we can halt the algorithm before arriving at a single cell, and the appropriate resolution may be chosen by the user or determined by other means.

## 5   Discussion

We have described a version of the watershed algorithm for segmenting a density function on a 3-manifold that uses persistence to counteract the over-segmentation, a common problem with the watershed paradigm. Many steps needed to turn this into a useful piece of software for medical data have been left unanswered. Should the algorithm be used on the raw data or will it be necessary to first apply a transformation of the data, such as a denoising, a smoothing, or a subsampling procedure? Three-dimensional medical images are usually given over a subset of $\mathbb{R}^3$, most often a cube sampled at an integer lattice of points (the centers of the voxels decomposing the cube). How severe are the artifact caused by the regular arrangement of the data points? Is it worth adding the voxel centers and use the Delaunay triangulation of the resulting body centered cubic lattice for the segmentation? It has better shaped tetrahedra than those triangulating the integer lattice and can still be treated implicitly, running the algorithm one vertex star at a time. The segmentation is defined by the two-dimensional triangulation consisting of the marked simplices. Will it be necessary to simplify this triangulation, eg. using edge-contractions running in parallel with the segmentation as described in [1]?

## References

1. D. ATTALI, D. COHEN-STEINER AND H. EDELSBRUNNER. Extraction and simplification of iso-surfaces in tandem. *In* "Proc. 3rd Eurographics Sympos. Geom. Process., 2005", 139–148.
2. S. BEUCHER. Watersheds of functions and picture segmentation. *In* "Proc. IEEE Intl. Conf. Acoustic, Speech, Signal Process, 1982", 1928–1931.
3. S. BEUCHER. The watershed transformation applied to image segmentation. *Scanning Microscopy Suppl.* **6** (1992), 299–314.
4. L. P. CLARKE, R. P. VELTHUIZEN, M. A. CAMACHO, J. J. HEINE, M. VAIDYANATHAN, L. O. HALL, R. W. THATCHER AND M. L. SILBIGER. MRI segmentation: methods and applications. *Magn. Reson. Imag.* **13** (1995), 343–368.
5. H. EDELSBRUNNER. *Geometry and Topology for Mesh Generation.* Cambridge Univ. Press, England, 2001.
6. H. EDELSBRUNNER. Surface tiling with differential topology. *In* "Proc. 3rd Eurographics Sympos. Geom. Process., 2005", 9–11.

7. H. EDELSBRUNNER, J. HARER, V. NATARAJAN AND V. PASCUCCI. Morse-Smale complexes for piecewise linear 3-manifolds. *In* "Proc. 19th Ann. Sympos. Comput. Geom., 2003", 361–370.

8. H. EDELSBRUNNER, D. LETSCHER AND A. ZOMORODIAN. Topological persistence and simplification. *Discrete Comput. Geom.* **28** (2002), 511–533.

9. F. FRIEDLANDER AND F. MEYER. A sequential algorithm for detecting watersheds on a gray level image. *Acta Stereol.* **6** (1987), 663–668.

10. R. M. HARALICK AND L. G. SHAPIRO. Image segmentation techniques. *Comput. Vis. Graph. Im. Proc.* **29** (1985), 100–132.

11. S. HAYKIN. *Neural Networks: a Comprehensive Foundation.* MacMillan College, New York, 1994.

12. S. L. HOROWITZ AND T. PAVLIDIS. Picture segmentation by a tree traversal algorithm. *J. Assoc. Comput. Mach.* **23** (1976), 368.

13. J. B. A. MAINTZ AND M. A. VIERGEVER. A survey of medical image registration. *Med. Im. Anal.* **2** (1998), 1–36.

14. I. N. MANOUSAKAS, P. E. UNDRILL, G. G. CAMERON AND T. W. REDPATH. Split-and-merge segmentation of magnetic resonance medical images: performance evaluation and extension to three dimensions. *Comput. Biomed. Res.* **31** (1998), 393–412.

15. Y. MATSUMOTO. *An Introduction to Morse Theory.* Translated from Japanese by K. Hudson and M. Saito, Amer. Math. Soc., 2002.

16. T. MCINERNEY AND D. TERZOPOULOS. Deformable models in medical image analysis: a survey. *Med. Im. Anal.* **1** (1996), 91–108.

17. J. MILNOR. *Morse Theory.* Princeton Univ. Press, New Jersey, 1963.

18. J. R. MUNKRES. *Elements of Algebraic Topology.* Addison-Wesley, Redwood City, California, 1984.

19. D. L. PHAM, C. XU AND J. L. PRINCE. Current methods in medical image segmentation. *Annu. Rev. Biomed. Engin.* (2000), 315–337.

20. J. ROERDINK AND A. MEIJSTER. The watershed transform: definitions, algorithms, and parallelization strategies. *Fundamenta Informaticae* **41** (2000), 187–228.

21. P. K. SAHOO, S. SOLTANI AND A. K. C. WONG. A survey of thresholding techniques. *Comput. Vision Graphics Im. Process.* **41** (1988), 233–260.

22. J. SIJBERS, P. SCHEUNDERS, M. VERHOYE, A. VAN DER LINDEN, D. VAN DYCK AND E. RAMAN. Watershed-based segmentation of 3D MR data for volume quantization. *Magn. Reson. Imag.* **15** (1997), 679–688.

23. J. S. SURI, K. LIU, S. SINGH, S. N. LAXMINARAYAN, X. ZENG AND L. REDEN. Shape recovery algorithms using level sets in 2-D/3-D medical imagery: a state-of-the-art review. *IEEE Trans. Inform. Techn. Biomed.* **6** (2002), 8–28.

24. R. THOM. Sur une partition en cellules associée à une fonction sur une variété. *Comptes Rendus l'Acad. Sci.* **228** (1949), 973–975.

25. L. VINCENT AND P. SOILLE. Watersheds of digital spaces: an efficient algorithm based on immersion simulations. *IEEE Trans. Pattern Anal. Mach. Intell.* **13** (1991), 583–593.

26. G. Z. YANG, P. BURGER, D. N. FIRMIN AND S. R. UNDERWOOD. Structure adaptive anisotropic image filtering. *Image Vision Comput.* **14** (1996), 135–145.

## Appendix

We now describe the actions taken in the general case, subsuming all types of vertices other than minima and maxima. Letting $u$ be such a vertex, the level set $f^{-1}(f(u))$ intersects its link in $j + 1 \geq 1$ closed curves decomposing the link into $j + 2$ open

regions. Using the reduced Betti numbers of the lower link, we count $\tilde{\beta}_0 + 1$ regions below $u$ and $\tilde{\beta}_1 + 1$ regions above $u$ so that $j = \tilde{\beta}_0 + \tilde{\beta}_1$. Each region below $u$ retracts to a component of the lower link, and each region above $u$ corresponds to a hole of a region below $u$ or to the outside. Instead of one tree or cactus, we construct $\tilde{\beta}_0 + 1$ multi-cacti with a total of $\tilde{\beta}_1$ cycles. Without counting components and holes, the right number of cacti and cycles arises as a side-effect of the process that adapts the family of cacti to the topology of the lower link.

Step 1.1. Construct a forest $T_u$ that contains a spanning tree within each component of the lower link, thus touching all vertices in the lower link.

Step 1.2. Thicken $T_u$ to a subcomplex $S_u$ of the lower link by iteratively adding triangles that already have two edges in $S_u$, until there are no more such triangles.

We may think of the thickening operation as a sequence of anti-collapses, each adding a triangle and its third edge. An anti-collapse does not change the homotopy type, which implies that the final complex $S_u$ generated in Step 1.2 consists of $\tilde{\beta}_0 + 1$ contractible components. We further expand these components until they cover the entire lower link. We do this by adding edges that change $T_u$ from a family of trees to a family of cacti.

Step 1.3. Initialize $C_u = T_u$ and iterate until $S_u$ is the lower link:

    Step 1.3.1. Add an edge of the lower link to $S_u$ and also to $C_u$.

    Step 1.3.2. Thicken $S_u$ by anti-collapsing triangles on both sides of the new edge until no further anti-collapses are possible.

We call the edges added in Step 1.3.1 *bridges* because they form cycles, turning trees into cacti. Note that the thickening process guarantees that every new cycle goes around a new hole in the lower link. To continue, we shrink $C_u$, retaining only a minimal set of edges needed to touch all stained vertices and preserve the topological type. Call a vertex *free* if it belongs to exactly one edge in $C_u$.

Step 1.4. Extract a family of cacti by iterating the following steps:

    Step 1.4.1. Find an unstained free vertex in $C_u$.

    Step 1.4.2. Collapse the incident edge by removing the vertex and the edge from $C_u$.

Steps 1.1 to 1.4 are illustrated in Figure 4. Given the input $C_u$ for Step 1.4, the result of the sequence of collapses is unique, unless there is a component without cycles and without stained vertices. The edge collapses reduce this component to a single but arbitrary vertex in this component. The remaining two steps use the simplices in $C_u$ to mark and the simplices in $S_u - C_u$ to label the simplices in the lower star.

Step 2. Mark the triangles and edges in the lower star whose faces in the lower link are the edges and vertices of $C_u$.

The family of cacti together with the marked edges connecting stained vertices to the outside decompose the subset of points in the link that are below $u$ into bays. Each bay touches the curves in which $f^{-1}(f(u))$ intersects the link and thus neighbors the complementary subset of points above $u$. In the last step, we visit the bays and label corresponding simplices in the lower star. Call an edge *free* if it belongs to exactly one triangle in $S_u$.
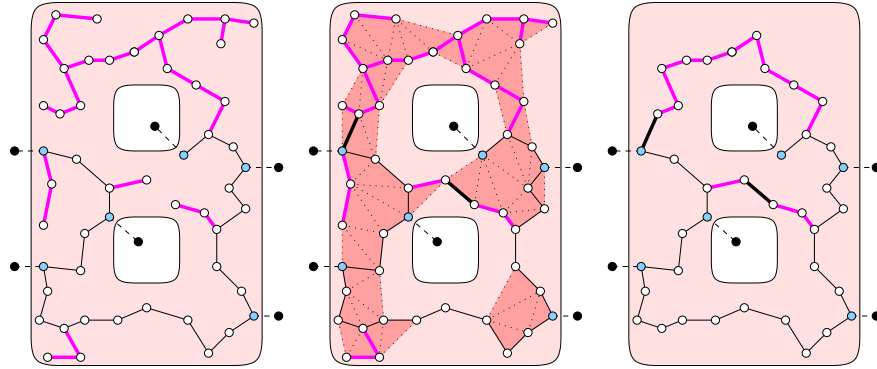
Fig. 4: Left: the level set of $u$ intersects the link in three curves forming a double-annulus with two holes and the outside region. We see a spanning tree that touches every vertex in the double-annulus. Middle: the tree is thickened by incrementally adding the shaded triangles. When this process halts, we add a (thick black) bridge, turning the tree into a cactus, and repeat until the entire lower link is covered. Right: the cactus that spans all stained vertices extracted from the thickened complex. It contains both bridges and is a deformation retract of the double-annulus.

Step 3. Shrink $S_u$ back to $C_u$ by iteratively collapsing free edges and vertices (the edge collapses undo the earlier anti-collapses and the vertex collapses repeat the collapses used to shrink the cacti):

  Step 3.1. Find a free edge or vertex, $v$ in $S_u - C_u$. In either case, $v$ belongs to a triangle in the link that does not belong to $S_u$. Let $\ell$ be the label of the corresponding tetrahedron in the star.

  Step 3.2. Let $\tau$ be the unique triangle or edge in $S_u$ that contains $v$. Label the simplices in the star that correspond to $v$ and $\tau$ with $\ell$.

  Step 3.3. Remove $v$ and $\tau$ from $S_u$.

As mentioned earlier, it is possible that a component of $C_u$ consists of a single unstained vertex, $v$. In this case, Step 3 labels all simplices in the corresponding component of the lower star, except for the edge connecting $u$ to $v$, which it marks. A special case arises if there are no other components in $C_u$, which is the regular case. Then $u$ itself is free and we collapse the edge by unmarking $u$ and $uv$ and labeling both consistent with the tetrahedra in their stars. As a side-effect, $v$ is no longer stained.