

Morse-Smale Complexes for Piecewise Linear 3-Manifolds ^{*}

Herbert Edelsbrunner[†], John Harer[‡], Vijay Natarajan[§] and Valerio Pascucci[¶]

Abstract

We define the Morse-Smale complex of a Morse function over a 3-manifold as the overlay of the descending and ascending manifolds of all critical points. In the generic case, its 3-dimensional cells are shaped like crystals and are separated by quadrangular faces. In this paper, we give a combinatorial algorithm for constructing such complexes for piecewise linear data.

Keywords. Computational geometry and topology, Morse theory, densities, triangulations, combinatorial algorithms.

1 Introduction

Morse functions are used by differential topologists to study the topology of manifolds [12, 13]. We use their results but pursue a different goal, namely that of studying topological features in natural phenomena.

Motivation. A three-dimensional Morse function is a generic smooth map from a 3-manifold to the real line. There is an abundance of natural phenomena that can be modeled by such functions. In oceanography, we study the distribution of temperature and other measurements over the Earth's oceans. In medical imaging, we reconstruct the inside of a living body from density distributions measured by MRI and other sensing technology. In x-ray crystallography, we determine the conformations of proteins and other molecules from

electron densities derived from x-ray diffractions. In each case, essential information is obtained from variations of the density over the space. Morse theory offers the basic mathematical language to reason qualitatively and quantitatively about this variation. In oceanography, we might be interested in the temperature extrema and how they change over time. In medical imaging, we use sharp changes in density to segment the body into bone, tissue and other constituents. In x-ray crystallography, we reconstruct geometric structure by following ridges connecting maxima in the electron density.

Related work. Three-dimensional densities are commonly visualized by drawing one or several level sets. In three-dimensional Euclidean space, such a set is generically a 2-manifold, often referred to as an iso-surface, which divides the space into inside and outside. The 1-parameter family of iso-surfaces sweeps out each cell in the Morse-Smale complex in a predictable manner, starting at the minimum and proceeding towards the opposite maximum while crossing the boundary everywhere at a right angle. The most popular method for computing an iso-surface is the marching cube algorithm, which assumes the density is given by its values at the vertices of a regular cubic grid [11]. Extensions and improvements of this algorithm can be found in [9, 23].

The marching cube algorithm visits the entire grid, which implies a running time proportional to the number of grid cells. A significant improvement in performance can be achieved by limiting the traversal to those cells that have a non-empty intersection with the constructed iso-surface. Starting at a 'seed edge', the algorithm traverses the cells following the component of the iso-surface as it is uncovered [2]. A minimal collection of seed edges that touches each component of every level set is provided by a minimal covering of the Reeb graph [17], stored for quick access in a hierarchical data structure referred to as the contour tree [22]. The Reeb graph is a compressed representation of the components, but it has no geometric information related to the gradient flow as expressed by the Morse-Smale complex. Extensions and improvements of the original algorithm for constructing contour trees can be found in [3, 16, 21].

Another concept related to Morse-Smale complexes is the

^{*}Research by the first and third authors is partially supported by NSF under grants EIA-99-72879 and CCR-00-86013. Research by the second author is partially supported by NSF under grant DMS-01-07621. This work was performed under the auspices of the U. S. Department of Energy by University of California Lawrence Livermore National Laboratory under contract No. W-7405-Eng-48.

[†]Departments of Computer Science and Mathematics, Duke University, Durham, North Carolina and Raindrop Geomagic, Research Triangle Park, North Carolina.

[‡]Department of Mathematics and Computer Science, Duke University, Durham, North Carolina.

[§]Department of Computer Science, Duke University, Durham, North Carolina.

[¶]Lawrence Livermore National Laboratory, Livermore, California.

medial axis of a shape in three-dimensional Euclidean space. As introduced by Blum [1], it is the set of centers of spheres that touch the boundary of the shape in at least two points without crossing it. Medial axes are used in a wide variety of applications, including shape representation [4, 18], mesh generation [19], geometric modeling [20], motion planning [10], image processing [15] and computer vision [24]. If the boundary is an orientable 2-manifold embedded in three-dimensional Euclidean space, we may define the signed distance as a function over the space. The medial axis then consists of arcs and quadrangles in the Morse-Smale complex.

Results. A fundamental difficulty in applying Morse theoretic ideas to scientific problems is the lack of smoothness in real data. Most commonly, information is gathered by point probes, and to turn these probes into a generic smooth function is a formidable task. We argue that the construction of such a function is also a questionable step if the goal is to compute and study topological features in the data, mostly because understanding the latter seems necessary to successfully do the former. Instead, we take a combinatorial approach and simulate smoothness to the extent necessary to make things work. The main results of this paper are combinatorial and algorithmic in nature:

- (i) the introduction of quasi Morse-Smale complexes as combinatorial analogs of the CW complexes defined by the descending and ascending manifolds of smooth functions;
- (ii) a combinatorial algorithm for constructing a quasi Morse-Smale complex with guaranteed structural correctness.

We believe that these results lay the ground-work for a large-scale application of Morse theoretic ideas to data sets in the sciences, engineering and medicine.

Outline. Sections 2 and 3 present the necessary background from Morse theory and combinatorial topology. Sections 4 to 7 describe the algorithm for constructing a quasi Morse-Smale complex for three-dimensional piecewise linear density data. Section 8 concludes the paper.

2 Smooth 3-Manifolds

In this section, we introduce the Morse theoretic concepts used in this paper. We refer to [12, 13] for further background.

Morse functions. Let \mathbb{M} be a smooth compact 3-manifold without boundary. Examples are the 3-sphere, which consists of all points at unit distance from the origin in \mathbb{R}^4 , and the 3-torus, which can be obtained by identifying opposite square faces of a three-dimensional cube. Let $f : \mathbb{M} \rightarrow \mathbb{R}$ be

a smooth map. The differential of f at a point $p \in \mathbb{M}$ is a linear map from the tangent space at p to \mathbb{R} , $df_p : T\mathbb{M}_p \rightarrow \mathbb{R}$. A point $p \in \mathbb{M}$ is *critical* if df_p is the zero map, otherwise it is *regular*. Given a local coordinate system, the *Hessian* at p is the matrix of second order partial derivatives:

$$H(p) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(p) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(p) & \frac{\partial^2 f}{\partial x_1 \partial x_3}(p) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(p) & \frac{\partial^2 f}{\partial x_2^2}(p) & \frac{\partial^2 f}{\partial x_2 \partial x_3}(p) \\ \frac{\partial^2 f}{\partial x_3 \partial x_1}(p) & \frac{\partial^2 f}{\partial x_3 \partial x_2}(p) & \frac{\partial^2 f}{\partial x_3^2}(p) \end{bmatrix}.$$

A critical point p is *non-degenerate* if the Hessian at p is non-singular. The function f is called a *Morse function* if all critical points are non-degenerate and $f(p) \neq f(q)$ whenever $p \neq q$ are critical. The Morse Lemma states that if p is non-degenerate we can choose local coordinates and signs such that

$$f(x_1, x_2, x_3) = f(p) \pm x_1^2 \pm x_2^2 \pm x_3^2$$

in a local neighborhood of p . Note this implies that non-degenerate critical points are isolated. The number of minuses is the *index* of the critical point. It is independent of the coordinate system and equals the number of negative eigenvalues of $H(p)$.

In three dimensions, there are four types of non-degenerate critical points: *minima* have index 0, *1-saddles* have index 1, *2-saddles* have index 2, and *maxima* have index 3. We get intuitive local pictures by drawing a small sphere around the point p . The level curve of points x with $f(x) = f(p)$ decomposes the sphere into *oceans*, consisting of points x with $f(x) < f(p)$, and *continents*, consisting of points x with $f(x) > f(p)$. Figure 1 shows the local pictures of a regular point and of the four types of non-degenerate critical points.

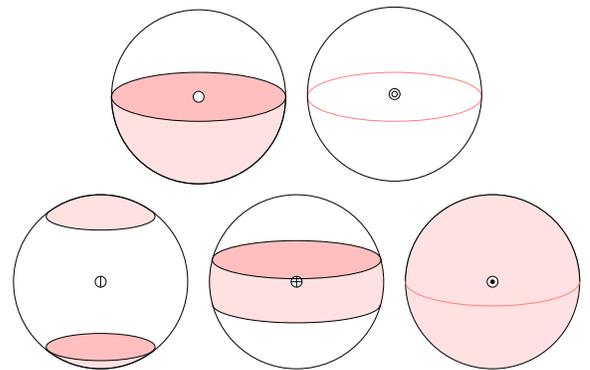


Figure 1: The local pictures with shaded oceans and white continents of a regular point, a minimum, a 1-saddle, a 2-saddle, and a maximum. Take notice of the symbols used to mark the different types of vertices at the centers of the spheres.

Descending and ascending manifolds. Given a Riemannian metric on \mathbb{M} and a local coordinate system with or-

thonormal tangent vectors $\frac{\partial}{\partial x_i}(p)$, the *gradient* of f at p is

$$\nabla f(p) = \left[\frac{\partial f}{\partial x_1}(p), \frac{\partial f}{\partial x_2}(p), \frac{\partial f}{\partial x_3}(p) \right]^T.$$

It is the zero vector iff p is critical. An *integral line* $\gamma : \mathbb{R} \rightarrow \mathbb{M}$ is a maximal path whose velocity vectors agree with the gradient: $\frac{\partial \gamma}{\partial s}(s) = \nabla f(\gamma(s))$ for all $s \in \mathbb{R}$. Each integral line is open at both ends, and we call $\text{org } \gamma = \lim_{s \rightarrow -\infty} \gamma(s)$ the *origin* and $\text{dest } \gamma = \lim_{s \rightarrow \infty} \gamma(s)$ the *destination* of γ . Both are necessarily critical points of f . Integral lines are pairwise disjoint. We consider each critical point as an integral line by itself, and with this stipulation the integral lines partition \mathbb{M} . We use them to decompose \mathbb{M} into regions of similar flow patterns. The *descending* and *ascending manifolds* of a critical point p are

$$\begin{aligned} D(p) &= \{p\} \cup \{x \in \mathbb{M} \mid x \in \text{im } \gamma, \text{dest } \gamma = p\}, \\ A(p) &= \{p\} \cup \{x \in \mathbb{M} \mid x \in \text{im } \gamma, \text{org } \gamma = p\}, \end{aligned}$$

where $\text{im } \gamma$ is the image of the path γ on \mathbb{M} . If x and y are points different from p that belong to the descending and the ascending manifolds of p then $f(x) < f(p) < f(y)$. This implies that $D(p) \cap A(p) = p$. The descending manifolds of f are the ascending manifolds of $-f$ and, symmetrically, the ascending manifolds of f are the descending manifolds of $-f$. This implies that the two types of manifolds have the same structural properties. Specifically, the descending manifold of a critical point p of index i is an open cell of dimension $\dim D(p) = i$. Since the integral lines partition \mathbb{M} , so do the descending manifolds. Moreover, they form a complex as the boundary of every cell is the union of lower-dimensional cells that are its *faces*. The ascending manifolds form a dual complex: for critical points p and q of f , $\dim D(p) = 3 - \dim A(p)$, and $D(p)$ is a face of $D(q)$ iff $A(q)$ is a face of $A(p)$.

Morse-Smale complexes. A Morse function f is *Morse-Smale* if the descending and ascending manifolds intersect only transversally. Suppose $D(p)$ and $A(q)$ have non-empty common intersection. If $\dim D(p) = 2$ and $\dim A(q) = 1$ then the transversality assumption implies $D(p) \cap A(q) = p = q$. In the more interesting case in which both are 2-manifolds, $A(p)$ and $D(q)$ are faces of $A(q)$ and $D(p)$ and, as illustrated in Figure 2, the common intersection is a simple path connecting the two critical points. Following [7], we define the cells of the *Morse-Smale complex* as the components of the sets $D(p) \cap A(q)$, over all critical points p and q of f . By definition, each cell of the Morse-Smale complex is a union of integral lines that all share the same origin q and the same destination p . The dimension of the cell is then the difference between the two indices. We call the cells of dimension 0 to 3 *nodes*, *arcs*, *quadrangles*, and *crystals*. Each two-dimensional cell is indeed a quadrangle, but its boundary may be glued to itself. The prototypical case of a crystal is a cube, which we imagine standing on its tip, but more interesting cases are possible.

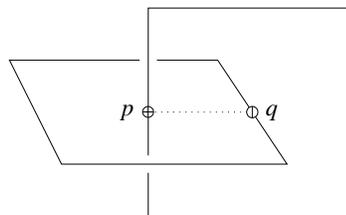


Figure 2: The dotted line is the common intersection of the descending 2-manifold of p and the ascending 2-manifold of q .

3 Piecewise Linear 3-Manifolds

We are interested in algorithms that work for piecewise linear functions obtained from point measurements. In this section, we introduce the necessary terminology, and we discuss some of the difficulties that arise when we transport concepts from the smooth to the piecewise linear category.

Triangulation. Let K be a simplicial complex that triangulates the 3-manifold \mathbb{M} . This means there is a homeomorphism between \mathbb{M} and the underlying space of K , but to simplify the discussion, we assume that \mathbb{M} is the underlying space. The complex consists of simplices of dimension 0 to 3, which we refer to as *vertices*, *edges*, *triangles* and *tetrahedra*. The *star* of a simplex σ consists of all simplices that contain σ as a face, including σ itself, and the *link* consists of all faces of simplices in the star that are disjoint from σ :

$$\begin{aligned} \text{St } \sigma &= \{\tau \in K \mid \sigma \subseteq \tau\}, \\ \text{Lk } \sigma &= \{v \in K \mid v \subseteq \tau \in \text{St } \sigma, v \cap \sigma = \emptyset\}. \end{aligned}$$

For example, if σ is a vertex then the link is a triangulation of the 2-sphere. Let $f : \mathbb{M} \rightarrow \mathbb{R}$ be a continuous function that is linear on every simplex of K . To say this more formally, we note that every point x in a simplex is a unique convex combination of its vertices u_ℓ : $x = \sum_\ell \lambda_\ell u_\ell$ with $1 = \sum_\ell \lambda_\ell$ and $\lambda_\ell \geq 0$ for all ℓ . Assuming f is given at the vertices, we have $f(x) = \sum_\ell \lambda_\ell f(u_\ell)$. We will refer to f as a height function and feel free to use relative terms such as ‘higher’ and ‘highest’. It will be convenient to assume that no two vertices have the same height, which can be justified computationally by simulating a perturbation, as described in [6, Section 1.4]. The *lower star* (*upper star*) of a vertex u contains all simplices in the star for which u is the highest (lowest) vertex, and the *lower link* (*upper link*) contains all simplices in the link that are faces of the lower star (upper star):

$$\begin{aligned} \text{St}_- u &= \{\tau \in \text{St } u \mid x \in \tau \implies f(x) \leq f(u)\}, \\ \text{St}^+ u &= \{\tau \in \text{St } u \mid x \in \tau \implies f(x) \geq f(u)\}, \\ \text{Lk}_- u &= \{v \in \text{Lk } u \mid v \subseteq \tau \in \text{St}_- u\}, \\ \text{Lk}^+ u &= \{v \in \text{Lk } u \mid v \subseteq \tau \in \text{St}^+ u\}. \end{aligned}$$

As in the smooth case, we draw the level curve of points x with $f(x) = f(u)$ to decompose the link into oceans and

continents. Each ocean retracts to a homotopy equivalent component of the lower link, and each continent retracts to a homotopy equivalent component of the upper link.

Critical vertices. Strictly speaking, critical points of f are not defined, but we may use small bump functions and think of f as the limit of a series of smooth maps. This is the intuition we use to transport concepts and results from the smooth to the piecewise linear category. We use lower links and their reduced Betti numbers to distinguish regular from critical vertices and to classify the latter. The reduced Betti numbers are denoted as $\tilde{\beta}_k$. They are the same as the common un-reduced Betti numbers, except that $\tilde{\beta}_0 = \beta_0 - 1$ for non-empty lower links, and $\tilde{\beta}_{-1} = 1$ for empty lower links [14]. Since lower links are two-dimensional, only $\tilde{\beta}_{-1}$ through $\tilde{\beta}_2$ can be non-zero. As shown in Table 1, the simple critical points are the ones that have exactly one non-zero reduced Betti number, which is equal to one. A *multiple saddle*

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

Table 1: The classification of regular and simple critical points using reduced Betti numbers.

is a vertex that falls outside the classification of Table 1 and therefore satisfies $\tilde{\beta}_{-1} = \tilde{\beta}_2 = 0$ and $\tilde{\beta}_0 + \tilde{\beta}_1 \geq 2$. It can be unfolded into simple 1-saddles and 2-saddles. One way to do that is to repeatedly cut the link along a circle that intersects the level curve separating the oceans and continents in exactly two points. The reduced Betti numbers on the two sides add up to the original ones: $\tilde{\beta}_k = \tilde{\beta}_{kL} + \tilde{\beta}_{kR}$, for $k = 0, 1$. We can always choose the circle such that the sum of the reduced Betti numbers are non-zero on both sides. It follows that the reduction ends after $\tilde{\beta}_0 + \tilde{\beta}_1 - 1$ cuts and generates $\tilde{\beta}_0$ 1-saddles and $\tilde{\beta}_1$ 2-saddles.

Quasi Morse-Smale complex. We construct a complex that is structurally indistinguishable from the Morse-Smale complex by taking open manifolds made up of simplices in K . It is a decomposition of space into crystals in which the boundary of each crystal is a quadrangulation. The function f has its critical points at the nodes of this complex and is monotonic within all the arcs, quadrangles and crystals. It differs from the Morse-Smale complex because the arcs and quadrangles may not be those of maximal ascent and descent. Let U, V, X and Y be the sets of minima, 1-saddles, 2-saddles and maxima of f , let R, S and T be the sets of arcs that connect minima to 1-saddles, 1-saddles to 2-saddles, and 2-saddles to maxima respectively, and let P and Q be the sets of quadrangles with nodes from U, V, X, V and V, X, Y, X

in that order, respectively, around the boundary. We define a *quasi Morse-Smale complex* of f as a decomposition of \mathbb{M} into open cells that satisfies the following properties:

- (i) all nodes are from $U \cup V \cup X \cup Y$, all arcs are from $R \cup S \cup T$, and all quadrangles are from $P \cup Q$,
- (ii) there are no critical points within the arcs, quadrangles and crystals, and
- (iii) each arc in S is on the boundary of four quadrangles, which in a cyclic order alternate between P and Q .

Note that a quasi Morse-Smale complex can be split into complexes defined by U, P and Y, Q . These are complexes that are structurally indistinguishable from those of the descending and ascending manifolds.

Simulating disjointness. Integral lines are not well defined for piecewise linear manifolds. So, following [7], we construct monotonic curves and surfaces that never cross. These curves and surfaces can merge together and fork later. When a curve or surface merges with another curve or surface, we pretend that they remain infinitesimally close to each other without crossing until they either fork or reach a common critical point.

4 Algorithm

In this section, we give an overview of the algorithm and describe some of the fundamental operations. Detailed descriptions of how we construct the descending and ascending manifolds will be given in Sections 5, 6 and 7.

Overview. A quasi Morse-Smale complex is constructed during two sweeps over the 3-manifold. The first sweep is in the order of decreasing function value or height and computes the descending manifolds. The second sweep is in the order of increasing height, which is the preferred order for computing the ascending manifolds. However, instead of computing the two collections independently, we use the structure provided by the descending manifolds and add the ascending manifolds accordingly.

Step 1. Construct the complex formed by the descending manifolds.

Step 2. Construct the ascending manifolds in pieces inside the cells formed by the descending manifolds.

Some routing decisions in Step 1 require rudimentary structural information about the ascending 2-manifolds, so we compute that already in Step 1. We compute the intersections between the descending and the ascending 2-manifolds before we construct the latter. It is in fact easier to compute these intersections first and then widen them into the ascending 2-manifolds. To streamline our description of the various steps in the algorithm, we denote the vertices of K by p_1, p_2, \dots, p_n assuming $f(p_1) > f(p_2) > \dots > f(p_n)$.

Links and critical vertices. We assume a data structure for the triangulation K of \mathbb{M} that connects neighboring simplices so that a local walk can be performed in constant time per visited simplex. An example of such a representation is the edge-facet data structure described in [5]. It stores ordered triangles linked into rings around shared edges. To illustrate the functionality of this data structure, consider the computation of the link of a vertex $p = p_i$. Letting puv be one of the triangles that share that vertex, we use depth-first search to traverse all triangles in the star. For each visited triangle pxy , the edge xy belongs to the link of p and so do the triangles that precede and succeed pxy in the ring around xy . Given the initial triangle puv , the search takes time proportional to the number of edges in the link.

With an additional test of the vertex heights, we can identify the lower link as a subcomplex of the link. As discussed in Section 3, we use the reduced Betti numbers of the lower link to classify the vertex p as regular, minimum, 1-saddle, 2-saddle, maximum or multiple saddle. We get the reduced Betti numbers by keeping track of the components in the lower link. If there are no components then $\tilde{\beta}_{-1} = 1$ and $\tilde{\beta}_k = 0$ for all $k \neq -1$, so p is a minimum. If the lower link is equal to the link then $\tilde{\beta}_2 = 1$ and $\tilde{\beta}_k = 0$ for all $k \neq 2$, so p is a maximum. Otherwise, $\tilde{\beta}_{-1} = \tilde{\beta}_2 = 0$ and $\tilde{\beta}_0$ is one less than the number of components. We get $\tilde{\beta}_1$ from $\tilde{\beta}_0$ and the Euler characteristic $\chi = s_0 - s_1 + s_2$, where s_k is the number of k -simplices in the lower link of p : $\tilde{\beta}_1 = \tilde{\beta}_0 + 1 - \chi$. According to Table 1, p is regular if $\tilde{\beta}_0 = \tilde{\beta}_1 = 0$ and it is a multiple saddle combining $\tilde{\beta}_0$ 1-saddles and $\tilde{\beta}_1$ 2-saddles, otherwise.

Running time. By choice of the data structure representing the triangulation K of the manifold, the link of p_i can be computed in time proportional to its size. Similarly, the classification of p_i , which reduces to counting the simplices and the components in the lower link, can be done in time proportional to that size. By definition, the size of the link is the number of simplices it contains, and because it is a two-dimensional sphere, this is $3t_i + 2$, where t_i is its number of triangles. Each triangle belongs to only two links, which implies that the total size of all vertex links is

$$\sum_{i=1}^n 3t_i + 2 = 3t/2 + 2n,$$

where n is the number of vertices and t is the number of triangles in K . As we will see later, the above time analysis applies to most steps taken by our algorithm. Indeed, we typically work inside a vertex link and compute simple substructures, such as shortest-path trees and circles separating oceans and continents from each other. We will see that with the assumption of unit length edges both tasks and miscellaneous others can be performed in time proportional to the size of the link and, in total, proportional to the size of K .

Besides computing vertex links, the algorithm constructs descending and ascending manifolds, which intersect to form

the quasi Morse-Smale complex. Even though these manifolds are made of simplices in K , their total size can exceed the size of K by any arbitrary amount. This is because the manifolds may fold onto themselves and onto each other. A simplex in K can therefore belong to several manifolds and it can belong several times to a single manifold. Whatever the situation, the time needed to add simplices to the description of the quasi Morse-Smale complex is only proportional to the total size of its description.

In summary, the running time of the algorithm is bounded from above by a constant times $n \log n$ (for sorting the vertices) plus the input size (for constructing and analyzing the vertex links) plus the output size (for describing the quasi Morse-Smale complex).

5 Descending Manifolds

We compute the descending 1- and 2-manifolds simultaneously during one sweep. To simplify the presentation, we first discuss them separately and restrict our attention to simple critical points.

Descending 1-manifolds. Each descending 1-manifold is an open interval that belongs to a 1-saddle $p = p_i$. It consists of two descending arcs and we call p the *root* of the 1-manifold and of its arcs. As illustrated in Figure 3, the 1-manifold descends from its root on both sides and, by simulation of the Morse-Smale condition, ends at minima of f . It is possible that the two arcs end at the same minimum, but because they do not contain that minimum, their union is still an open interval and not a closed circle. In the Morse-

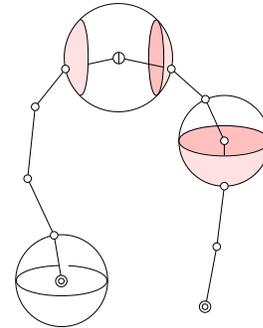


Figure 3: The descending 1-manifold rooted at a 1-saddle. The spheres sketch the links of the root, a regular point, and one of the two minima.

Smale case, all vertices of the 1-manifold except for its root are regular, but in the piecewise linear case it is also possible that the 1-manifold passes through a 2-saddle or 1-saddle p_j . We have $j > i$ because p_j is necessarily lower than the root. For an arc it makes little difference whether it passes through a regular or a critical point. However, since p_j starts its own descending manifold, we need to make sure that the

arcs descending from p_i and p_j are consistent in the sense of *simulated disjointness*. In most cases, this consistency will be automatic because we extend each arc by adding the edge from the current endpoint to the lowest vertex in its lower link. This choice of extension implies, for example, that once two arcs merge, they go together until they both end at the same minimum.

We distinguish between three operations in the construction of the descending 1-manifolds: starting, expanding and gluing. The same three operations also occur in the construction of descending 2-manifolds, and they are processed within the same logical structure. The starting operation applies if p is a 1-saddle and starts the two arcs of the corresponding 1-manifold using edges from p to the lowest vertex in each ocean of the link. The expanding operation continues all descending arcs ending at p by adding an edge from p to the lowest vertex in its lower link. An exception to this rule occurs if p is a 1-saddle. In this case, we (will later) start an ascending 2-manifold and we extend each descending arc to the lowest vertex in the specific ocean that is not separated from the arc by that 2-manifold. The gluing operation applies if p is a minimum, which it declares a node of the Morse-Smale complex, and glues the descending arcs ending at p to each other.

Structure of a 2-manifold. The construction of the descending 2-manifolds is considerably more complicated than that of 1-manifolds. We begin by discussing their structure and by formulating an invariant maintained by the algorithm. Each descending 2-manifold is an open disk that belongs to a 2-saddle, which we call its *root*. The disk descends from the root, which is its highest vertex. Its boundary is a circle consisting of descending 1-manifolds that meet at shared minima. The circle might be partially glued to itself along one or more arcs. Note that this is fundamentally different from the case in which the disk folds onto itself: the folding can be simulated away since it does not happen for smooth functions, while the boundary gluing is an inherent feature of descending 2-manifolds. It is important that the descending 2-manifold does not contain its boundary, else it would not necessarily be a disk. In the most extreme case, the boundary circle is a single vertex so that the closure of the disk is a sphere. This gives the disk the appearance of a pouch.

Beyond being an open disk which descends from its root, we require that the restriction of f to the descending 2-manifold has no critical points other than the maximum at its root $p = p_i$. This property is guaranteed by an invariant maintained during the construction. At any moment, we have an open disk whose boundary is partially final or *frozen* and partially *unfrozen*. The frozen boundary grows from the empty set to a collection of open segments, which eventually merge to form a complete circle. The unfrozen boundary shrinks from a complete circle to a collection of closed segments, until it eventually disappears.

DISK INVARIANT. Let q be a vertex in the unfrozen portion

of the boundary of a disk and let qu be an interior edge. Then u is either an interior vertex or a frozen boundary vertex, and $f(u) > f(q)$.

Note that the Disk Invariant prohibits interior edges that connect two unfrozen boundary vertices. This implies that as long as the entire boundary is unfrozen, there are no interior edges connecting two boundary vertices, and all edges descend from the interior to the boundary. Figure 4 illustrates the resulting structure of a descending disk. A regular vertex u in the restriction of f to the disk is characterized by a non-empty connected lower link. In other words, the edges in the star change between descending from u to descending towards u exactly twice around u . The disk is extended at

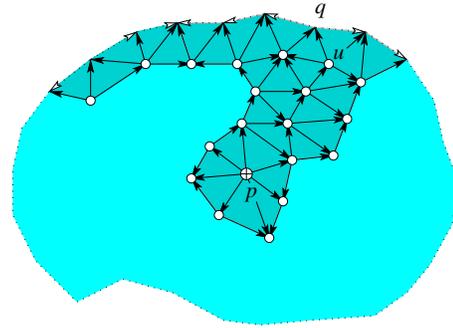


Figure 4: A portion of the triangulation of a partially constructed descending 2-manifold. The edges are oriented from the higher to the lower endpoints.

the highest unfrozen vertex q ; it either lies in the interior of an unfrozen boundary segment or is the endpoint of a frozen boundary segment. In the former case, all interior edges descend towards q . We maintain the Disk Invariant by extending the disk such that all newly added edges descend from q . It follows that the only new interior vertex, which is q itself, is a regular point of f restricted to the disk. In the latter case, we maintain the Disk Invariant by again extending the disk such that all newly added edges descend from q .

Starting a 2-manifold. We start a descending disk at every 2-saddle, and we extend descending disks at all unfrozen boundary vertices. Let $p = p_i$ be a 2-saddle, as shown in Figure 5, and let q be the lowest vertex in its link. By assumption, the lower link is a retract of the belt-like ocean around the link, and q belongs to that ocean. We start the corresponding descending disk by constructing a circle in the lower link, making sure that circle contains q as one of its vertices. Even though we call it a circle, it may fold onto itself, and sometimes such folding is unavoidable. There are many ways to construct such a circle. Our particular algorithm finds a shortest such circle using the shortest-path tree from q that spans the lower link. Assuming unit edge lengths, such a circle minimizes the number of edges. After constructing the tree, we classify non-tree edges in the lower

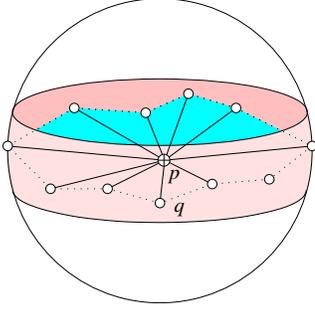


Figure 5: The disk rooted at p starts by connecting p to a circle in the belt-like ocean that passes through the lowest vertex q .

link depending on whether or not they separate the two continents. The circle is then defined by the separating non-tree edge in the lower link whose two endpoints minimize the sum of distances to q . Returning to the classification, we note that the tree cuts the link open but keeps it connected. If we cut along a non-tree edge, we split the link into two disks. If the edge does not separate then one of the disks contains both continents while the other is contained inside the ocean. The latter disk is triangulated and, by construction, its triangulation has all vertices on the boundary. We can therefore remove the triangles from the disks by repeated collapsing: at each step remove a triangle that has both edges on the boundary and declare the third edge a new boundary edge. The classification of non-tree edges in the lower link thus proceeds by repeated collapsing, which marks all non-separating edges and leaves all separating edges unmarked.

Expanding a 2-manifold. The interior vertices of a disk are typically regular points of f , although they can also be 1-saddles and 2-saddles. We first consider a regular point $p = p_i$ and assume it belongs to the boundary of a descending disk. Since we visit the vertices in the order of decreasing height, p is the highest boundary vertex adjacent to at least one unfrozen boundary edge. Figure 6 illustrates the two possible cases: one in which there are two neighbors, a and b , connected to p by unfrozen boundary edges, and the other in which there is only one such neighbor, c . The algorithm treats both cases similarly and simultaneously. Specifically, it constructs a shortest-path tree from the lowest vertex q in the lower link of p . The points a and b belong to the lower link and are therefore vertices of the tree. We connect a to q along the unique path in the tree and extend the corresponding disk by connecting p to the edges of that path. We do the same for b and for all other vertices that are connected to p by unfrozen boundary edges. It is possible that some paths fold onto each other or themselves, and we must keep track of sidedness as before.

There is no essential difference in the computations if p is a 2-saddle, except that p itself starts an additional descending disk. By using the same tree for starting disks and for

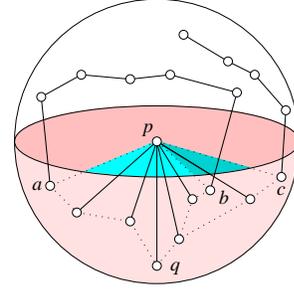


Figure 6: Two descending disks that touch p and intersect the link in a path each. One path starts and ends in the ocean while the other starts in the continent and ends in the ocean.

expanding disks we avoid intersections, but as usual, folding on themselves or each other is allowed. The case of a 1-saddle p can be more interesting. If the two neighbors of p along the boundary of the disk belong to opposite polar oceans in the link then we do the same computations within both oceans. The point p remains on the boundary, but its two neighbors change to the vertices that are adjacent along the descending 1-manifold rooted at p . Before continuing, we declare p and the two incident boundary edges frozen for the descending disk.

6 Simultaneous Construction

As mentioned earlier, the descending arcs and disks are really constructed simultaneously, in a single sweep over the 3-manifold. To get a flavor of how this is done, we discuss a multiple saddle $p = p_i$ characterized by $\tilde{\beta}_0 + \tilde{\beta}_1 \geq 2$. Its link has $\tilde{\beta}_0 + 1$ oceans and $\tilde{\beta}_1 + 1$ continents. We process p in five steps:

- Step 1.1. Start $\tilde{\beta}_1$ descending disks.
- Step 1.2. Prepare $\tilde{\beta}_0$ ascending disks.
- Step 1.3. Extend descending disks that touch p .
- Step 1.4. Start $\tilde{\beta}_0$ descending 1-manifolds.
- Step 1.5. Extend descending arcs that touch p .

The main difficulty is the coordination of the descending and ascending discs and arcs in such a way that they all intersect in a locally and globally consistent manner.

Families of circles. In Steps 1.1 and 1.2, we start one family of disks and prepare the starting of a second family. Each disk intersects the link of p in a circle, so we need two families of circles, one for the descending and the other for the ascending disks. The former are contained in the oceans and separate the continents, while the latter lie on the continents and separate the oceans, as illustrated in Figure 7. We extend the algorithm described in Section 5 to construct the first family of circles. As before, we begin with the shortest-path

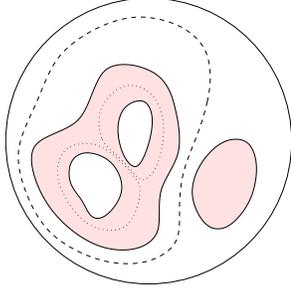


Figure 7: We draw $\tilde{\beta}_1 = 2$ (dotted) circles to separate the three continents and $\tilde{\beta}_0 = 1$ (dashed) circle to separate the two oceans. The descending disks that start at p intersect the link in the dotted circles, and the ascending disk intersects the link in the dashed circle.

tree from the lowest vertex q in every component of the lower link, and we classify non-tree edges in the lower link depending on whether or not they separate the continents into two non-empty sets. Once we have selected a separating edge, we add it to the tree of its endpoints (which now is a graph with one cycle), and we continue using collapses to eliminate edges that separate the continents in the same way. We repeat until we added $\tilde{\beta}_1$ edges to the collection of trees. These edges define the $\tilde{\beta}_1$ circles required in Step 1.1. We then repeat the same algorithm in the upper link of p , thus switching the roles of oceans and continents. This gives the $\tilde{\beta}_0$ circles required in Step 1.2. We note, however, that the construction of the second family is complicated by the presence of paths at which descending disks started at earlier vertices intersect the link. We next describe these complications and how we cope with them.

Transversal intersections. At the time we start descending disks and prepare ascending disks all rooted at $p = p_i$, we already have information on descending disks rooted at vertices p_j , with $j < i$. Among these disks, the ones that pass through p influence the construction of the circle families. Circles that start descending disks are unproblematic because the use of the shortest-path tree, both for starting and for expanding, implies non-crossing descending disks. We need some precautions to make sure that the prepared ascending disks either do not cross the descending disks or cross them transversally. In particular, if a descending disk meets the link of p in two different oceans then that disk has got to cross every ascending disk started by a circle that separates the two oceans. We cope with this difficulty by modifying the continents before drawing the circles. Each relevant descending disk meets the link of p in a path connecting the two vertices adjacent to p along its boundary. The Disk Invariant maintained for the descending disk implies that this path intersects at most two oceans and one continent. Clip the path to within that continent and let x and y be the endpoints of the clipped path. We distinguish between four cases:

- (i) both x and y lie in the continent (e.g. d_1 in Figure 8);
- (ii) x lies in the continent and y lies on the boundary of the continent (d_2, d_3 and d_4);
- (iii) x and y lie on a common boundary component (d_5);
- (iv) x and y lie on different boundary components (d_6).

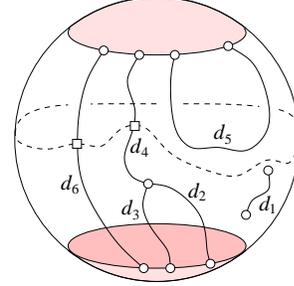


Figure 8: Some of the descending disks passing through p form barriers in our effort to draw circles preparing ascending disks within the continents. The squares are gateways at which the dashed circle may cross the paths.

In each case, we cut the continent open along the interior of the path from x to y . In other words, we form a barrier that prohibits a circle in the second family from crossing the path. Technically, we create this barrier by duplicating each edge and each vertex in the interior of the path. The two copies of a duplicated edge or vertex lie on different sides of the barrier and are connected to the simplices in the star that lie on the same side. When descending disks share a common endpoint in the continent. An example of this situation is the vertex shared by the paths d_2, d_3 and d_4 in Figure 8. Letting k be the number of such paths, the neighborhood of the shared vertex is cut into k wedges. We replace the vertex by k copies, one connected to each wedge. A path of type (iv) would prohibit any circle going around the continent as required. We thus designate the highest interior vertex as a *gateway* at which a circle may cross the path. We cut the continent along the rest of the path but not at the gateway. A similar situation arises when paths of type (ii) connect at shared endpoints and collectively cut a continent all the way from one ocean to another. Consider a connected component of such paths. It touches at least two oceans and we single out the one whose lowest vertex is the lowest among them. On every path in the component that connects to an ocean, other than the one we singled out, we designate the highest vertex as a gateway. After modifying the continents as described, we construct the shortest-path trees and rout the circles as explained before.

Descending arcs and spikes. For each ascending disk we start a (dual) descending 1-manifold that crosses the disk at p . Instead of determining the two oceans separated by that ascending disk and by no others, we start a descending arc by

connecting p to the lowest vertex in each ocean and thereafter pair the arcs to form the descending 1-manifolds.

The starting circles of ascending disks are also used in an essential way when we extend descending arcs that pass through p . Each such arc enters p from a continent and we extend it into the unique ocean that is not separated by any circle from the entry point. A complication arises when two or more descending disks share a common descending arc, such as the disks that meet the link in paths d_2 , d_3 and d_4 in Figure 8. Let x be the vertex at which the arc meets the link of p and consider a disk that meets the link in a path connecting x to a vertex y on the boundary of some ocean. In the easy case, the arc gets extended into the same ocean, but if the vertices y of these descending disks lie on two or more boundary components, then this cannot be the case for all of them. To resolve the apparent conflict, we attach a spike to each descending disk that expands into an ocean separated from that of the arc. This spike is an infinitesimally thin strip of the descending 2-manifold that is squeezed between two infinitesimally close descending 1-manifolds in its boundary. Note that the disks that get spikes are exactly the ones whose paths in the link receive gateways.

7 Ascending Manifolds

The construction of the ascending manifolds is similar to that of the descending manifolds, except for the complications caused by the fact that the latter already exist. The added constraints are expressed in terms of barriers formed within vertex links. We construct the ascending manifolds during a sweep of the 3-manifold in the direction of increasing function value. After computing the intersection curves between the descending and the ascending 2-manifolds and adding the ascending arcs connecting the 2-saddles with the maxima, we fill in the ascending 2-manifolds one quadrangle at a time.

Intersection curves and ascending arcs. Recall that for a Morse-Smale function on a 3-manifold, the intersection between a descending 2-manifold D and an ascending 2-manifold A is either empty or a curve connecting their two roots. From D 's point of view, the curve starts at a 1-saddle on its boundary and monotonically increases until it ends at its root. The Disk Invariant maintained during the construction of the descending disks implies that the restriction of f to D has no critical points other than the maximum at its root. To construct the curve, we thus start at the 1-saddle and repeatedly extend the path by connecting its endpoint to the highest adjacent vertex in the triangulation of D . The curves started at the various 1-saddles in the boundary may meet but they never cross and eventually all end at the root of D . Two curves in different descending 2-manifolds may also meet, but this intersection will be resolved when the descending 2-manifolds get resolved by simulation of an infinitesimal separation.

Next we construct the ascending 1-manifolds. Specifically, we start the two arcs of an ascending 1-manifold at every 2-saddle. The algorithm is similar to the one for descending 1-manifolds, except that now the primary concern in starting and expanding a 1-manifold is to avoid crossing any of the already established descending 2-manifolds. To understand these constraints consider the components of the link of a vertex p that are cut out by the descending 2-manifolds passing through p . Call these components the *slabs* of p and their intersections with the oceans and the continents the *lower* and *upper slabs*, respectively. When we start the ascending arcs we connect p with the highest vertices in the two upper slabs that are separated by the descending disk rooted at p but by none of the other descending disks that happen to pass through p . When we extend an ascending arc at p , we add the edge connecting p to the highest vertex in the slab from which the arc approached p .

Ascending disks. The intersection curves and ascending arcs decompose the ascending 2-manifolds into quadrangles. The lowest point of a quadrangle is the 1-saddle at which the 2-manifold is rooted. This 1-saddle is connected to 2-saddles by two continuous intersection curves emanating from the 1-saddle, and the 2-saddles are connected to a common maximum along ascending arcs. We now construct the individual quadrangles, which then fit together to form the ascending 2-manifold. Each quadrangle is constructed in a process similar to that for the descending disks. In this case, the frozen part of the boundary occurs when the boundary of the quadrangle meets either an intersecting curve or an ascending arc. Edges and vertices on these curves and arcs are frozen, except for the vertices where we transition from frozen to unfrozen edges, which are considered unfrozen. The process also maintains a property similar to the Disk Invariant with the inequality reversed.

The starting of an ascending quadrangle has already been prepared in the descending step. Let $p = p_i$ be a 1-saddle. The ascending disk at p meets the continent in a circle, this circle is cut into a collection of segments by the descending disks that pass through p , and the cone of p over each segment is the initial portion of a quadrangle. The endpoints of these segments lie on intersection curves. To discuss the expansion of an ascending quadrangle, suppose that p is the lowest point on its unfrozen boundary. In this case, the picture is dual to that of Figure 6: the two adjacent points on the boundary of the quadrangle either both lie in the continent of p or one lies in the continent and the other is frozen in the ocean. These points all lie in a single slab. In the first case, connect both points to the highest point in the slab that contains them, using the same algorithm that we used in the descending case. In the latter case, connect the single point to the point of the slab where the ascending arc or intersection curve emerges in the continent. Clearly, for this to be possible, we have to choose the intersection curves and ascending arcs carefully. We discuss this next.

Simultaneous construction. As in the descending case, we may actually construct intersection curves, ascending arcs and ascending quadrangles all in a single pass from bottom to top without resolving multiple saddles into simple ones. To see how this goes, consider a multiple saddle $p = p_i$ with $\tilde{\beta}_0 + 1$ oceans and $\tilde{\beta}_1 + 1$ continents. We process p in six steps:

- Step 2.1. Start $\tilde{\beta}_1$ ascending 1-manifolds.
- Step 2.2. Start intersection curves.
- Step 2.3. Start $\tilde{\beta}_0$ ascending disks.
- Step 2.4. Extend ascending arcs that touch p .
- Step 2.5. Extend intersection curves that touch p .
- Step 2.6. Extend ascending quadrangles that touch p .

In Step 2.1, we choose one arc for each continent and pair these ensuring that each pair of arcs is dual to a unique descending disk that we started at p . In Step 2.2, we start ascending intersection curves by connecting p to the gateways in its link. Due to paths and circles folding onto each other, several paths may cross several circles at one and the same gateway, and we start an intersection curve for each pair. Two contiguous gateways along a circle delimit the starting segment of an ascending quadrangle. In Step 2.3, we start the quadrangle by forming the cone of p over the segment. To do Step 2.4, we note that each ascending arc that passes through p enters the lower link in either a single slab or on the boundary between two slabs, which is necessarily part of a descending disk. We continue the arc by choosing the highest point in the corresponding slab or the highest point in the boundary between the two slabs. In the former case, we maintain simulated disjointness from the descending disks and in the latter we maintain tangency. To do Step 2.5, we continue the intersection curve by choosing the highest point in the upper link of p that lies on the correct descending disk. Step 2.6 is now clear, following the procedure for expanding an ascending quadrangle described above.

8 Discussion

This paper introduces the Morse-Smale complex for a function over a 3-manifold as a decomposition of the 3-manifold into crystals with quadrangular faces. It also gives an algorithm to construct a quasi Morse-Smale complex for a piecewise linear function that guarantees structural correctness. Letting n be the number of vertices in the input triangulation, the running time is proportional to $n \log n$ plus the size of the input triangulation plus the total size of the output manifolds describing the quasi Morse-Smale complex.

Many interesting issues still remain open. We can transform the quasi Morse-Smale complex into the Morse-Smale complex by applying a sequence of operations called handle slides. As described for 2-manifolds in [7], using this

approach we obtain a Morse-Smale complex that is numerically as accurate as the local rerouting operations used to control handle slides. For 3-manifolds, it is unclear how to find and order the handle slides that bring us closer to the Morse-Smale complex. It is also useful to have a hierarchical representation of the Morse-Smale complex while working with large data sets. We can create such a hierarchy by performing a sequence of cancellations of pairs of critical points ordered by persistence [8]. The details of this simplification process as applied to a quasi Morse-Smale complex still have to be investigated.

References

- [1] H. BLUM. *Models for the Perception of Speech and Visual Form*. MIT Press, Cambridge, 1967, 362–380.
- [2] C. L. BAJAJ, V. PASCUCCI AND D. SCHIKORE. Fast isocontouring for improved interactivity. In “Proc. IEEE Sympos. Vol. Viz., 1996”, 39–46.
- [3] H. CARR, J. SNOEYINK AND U. AXEN. Computing contour trees in all dimensions. In “Proc. 11th Ann. SIAM-ACM Sympos. Discrete Algorithms, 2000”, 918–926.
- [4] T. CULVER, J. KEYSER AND D. MANOCHA. Accurate computation of the medial axis of a polyhedron. In “Proc. ACM Sympos. Solid Model. Appl., 1999”, 179–190.
- [5] D. P. DOBKIN AND M. J. LASZLO. Primitives for the manipulation of three-dimensional subdivisions. *Algorithmica* **4** (1989), 3–32.
- [6] H. EDELSBRUNNER. *Geometry and Topology for Mesh Generation*. Cambridge Univ. Press, England, 2001.
- [7] H. EDELSBRUNNER, J. HARER AND A. ZOMORODIAN. Hierarchical Morse-Smale complexes for piecewise linear 2-manifolds. *Discrete Comput. Geom.*, to appear.
- [8] H. EDELSBRUNNER, D. LETSCHER AND A. ZOMORODIAN. Topological persistence and simplification. *Discrete Comput. Geom.* **28** (2002), 511–533.
- [9] A. GUÉZIEC AND R. HUMMEL. Exploiting triangulated surface extraction using tetrahedral decomposition. *IEEE Trans. Visualization Comput. Graphics* **1** (1995), 328–342.
- [10] L. GUIBAS, R. HOLLEMAN AND L. E. KAVRAKI. A probabilistic roadmap planner for flexible objects with a workspace medial axis based sampling approach. In “Proc. IEEE/RSJ Intl. Conf. Intell. Robots Systems, 1999”, 254–260.
- [11] W. E. LORENSEN AND H. E. CLINE. Marching cubes: a high resolution 3D surface construction algorithm. *Comput. Graphics* **21**, Proc. SIGGRAPH 1987, 163–169.
- [12] Y. MATSUMOTO. *An Introduction to Morse Theory*. Translated from Japanese by K. Hudson and M. Saito, Amer. Math. Soc., 2002.
- [13] J. MILNOR. *Morse Theory*. Princeton Univ. Press, New Jersey, 1963.

- [14] J. R. MUNKRES. *Elements of Algebraic Topology*. Addison-Wesley, Redwood City, California, 1984.
- [15] R. L. OGNIWICZ. Skeleton-space: A multi-scale shape description combining region and boundary information. In “Proc. Comput. Vision Pattern Recogn., 1994”, 746–751.
- [16] V. PASCUCCI AND K. COLE-MCLAUGHLIN. Efficient computation of the topology of level sets. *Algorithmica*, to appear.
- [17] G. REEB. Sur les points singuliers d’une forme de Pfaff complètement intégrable ou d’une fonction numérique. *Comptes Rendus de L’Académie ses Séances, Paris* **222** (1946), 847–849.
- [18] D. SHEEHY, C. ARMSTRONG AND D. ROBINSON. Shape description by medial axis construction. *IEEE Trans. Visualization Comput. Graphics* **2** (1996), 62–72.
- [19] A. SHEFFER, M. ETZION, A. RAPPOPORT AND M. BERCOVIER. Hexahedral mesh generation using the embedded Voronoi graph. *Engineering Comput.* **15** (1999), 248–262.
- [20] D. STORTI, G. TURKIYYAH, M. GANTER, C. LIM AND D. STAL. Skeleton-based modeling operations on solids. In “Proc. ACM Sympos. Solid Model. Appl., 1997”, 141–154.
- [21] S. TARASOV AND M. N. VYALI. Construction of contour trees in 3D in $O(n \log n)$ steps. In “Proc. 14th Ann. Sympos. Comput. Geom., 1998”, 68–75.
- [22] M. VAN KREVELD, R. VAN OOSTRUM, C. BAJAJ, V. PASCUCCI AND D. SCHIKORE. Contour trees and small seed sets for iso-surface traversal. In “Proc. 13th Ann. Sympos. Comput. Geom., 1997”, 212–220.
- [23] J. WILHELMS AND A. VAN GELDER. Topological consideration in iso-surface generation. *ACM Trans. Comput. Graphics* **24** (1990), 57–62.
- [24] A. Y. WU, S. K. BHASKAR AND A. ROSENFELD. Computation of geometric properties from the medial axis transform in $O(n \lg n)$ time. *Comput. Vision, Graphics, Image Process.* **34** (1986), 76–92.