Discrete Comput Geom 30:87–107 (2003)
DOI: 10.1007/s00454-003-2926-5

Discrete & Computational
Geometry

Hierarchical Morse–Smale Complexes for Piecewise Linear 2-Manifolds*

Herbert Edelsbrunner,[1] John Harer,[2] and Afra Zomorodian[3]

[1]Department of Computer Science, Duke University,
Durham, NC 27708, USA
edels@cs.duke.edu
and
Raindrop Geomagic,
Research Triangle Park, NC 27709, USA

[2]Department of Mathematics, Duke University,
Durham, NC 27708, USA
harer@math.duke.edu

[3]Department of Computer Science, University of Illinois,
Urbana, IL 61801, USA
afra@cs.stanford.edu

Abstract. We present algorithms for constructing a hierarchy of increasingly coarse Morse–Smale complexes that decompose a piecewise linear 2-manifold. While these complexes are defined only in the smooth category, we extend the construction to the piecewise linear category by ensuring structural integrity and simulating differentiability. We then simplify Morse–Smale complexes by canceling pairs of critical points in order of increasing persistence.

1. Introduction

In this paper we define the Morse–Smale complex decomposing a piecewise linear 2-manifold and present algorithms for constructing and simplifying this complex.

* Research by the first and third authors was partially supported by ARO under Grant DAAG55-98-1-0177. Research by the first author was also partially supported by NSF under Grants CCR-97-12088, EIA-9972879, and CCR-00-86013.

Motivation. Physical simulation problems often start with a space and measurements over this space. If the measurements are scalar values, we talk about a height function over that space. We use this name throughout the paper, although the functions can be arbitrary and do not necessarily measure height. Two-dimensional examples of height functions include intensity values of an image and the elevation of a terrain as parametrized by longitude and latitude. Three-dimensional examples include the temperature within a room and the electron density over a crystallized molecule. In all these examples, we seek to derive structures that enhance our understanding of the measurements.

Consider a geographic landscape modeled as a height function $h\colon D \to \mathbb{R}$ over a two-dimensional domain D. We can visualize h by a discrete set of iso-lines $h^{-1}(c)$, for constant height values c. The topological structure of iso-lines is partially captured by the contour tree [3], [4], [17]. If h is differentiable, we may define the gradient field consisting of vectors in the direction of the steepest ascent. Researchers in visualization have studied this vector field for some time [1], [5], [16]. The Morse–Smale complex captures the characteristics of this vector field by decomposing the manifold into cells of uniform flow. As such, the Morse–Smale complex represents a full analysis of the behavior of the vector field.

Often, however, the smooth domain D is sampled. No matter how dense the sampling, we encounter two critical issues: our theoretical notions, based on smooth structures, are no longer valid, and we have to distinguish between noise and features in the sampled data. Our goal in this work is resolve both issues in the piecewise linear (PL) domain.

Methods and Results. To extend smooth notions to PL manifolds, we use differential structures to guide our computations. We call this method the *simulation of differentiability* or *SoD* paradigm. Using SoD, we first guarantee the computed complexes have the same structural form as those in the smooth case. We then achieve numerical accuracy by means of transformations that maintain this structural integrity. The separation of combinatorial and numerical aspects of computation is similar to many algorithms in computational geometry. It is also the hallmark of the SoD paradigm. Our results are:

(i) an algorithm for constructing a complex whose combinatorial form matches that of the Morse–Smale complex,
(ii) an algorithm for deriving the Morse–Smale complex from the complex in (i) via local transformations,
(iii) an algorithm for constructing a hierarchy of Morse–Smale complexes, again via local transformations, and
(iv) the application of the algorithms to geographic terrain data.

Because of the theoretical nature of our endeavor, we devote most of our effort in this paper to (i)–(iii). While we include only a short section on (iv), we view this paper as the foundation for creating robust software for the scientific and engineering fields.

Outline. The rest of the paper is organized as follows. In Section 2 we introduce the theoretical background from Morse theory on smooth 2-manifolds. In Section 3 we extend these notions to PL domains, and discuss difficulties resulting from the absence of smoothness. Having computed a structurally correct complex in Section 4, we compute the Morse–Smale complex via transformations in Section 5. We introduce topological

persistence in Section 6, and use it to create a hierarchy of complexes in Section 7. We give some experimental results for geographic landscapes in Section 8, concluding the paper in Section 9.

2. Smooth 2-Manifolds

In this section we introduce concepts from Morse theory we need as the theoretical background for our work. We refer to [11] and [18] for further background.

Morse Functions. Let $\mathbb{M}$ be a smooth, compact 2-manifold without boundary and let $h\colon \mathbb{M} \to \mathbb{R}$ be a smooth map. The differential of h at the point a is a linear map $\mathrm{d}h_a\colon \mathrm{T}\mathbb{M}_a \to \mathrm{T}\mathbb{R}_{h(a)}$ mapping the tangent space of $\mathbb{M}$ at a to that of $\mathbb{R}$ at $h(a)$. (The tangent space of $\mathbb{R}$ at a point is simply $\mathbb{R}$ again, with the origin shifted to that point.) A point a is called *critical* if the map $\mathrm{d}h_a$ is the zero map. Otherwise, it is a *regular* point. At a critical point a we compute in local coordinates the *Hessian* of h,

$$H(a) = \begin{bmatrix} \dfrac{\partial^2 h}{\partial x^2}(a) & \dfrac{\partial^2 h}{\partial y\,\partial x}(a) \\ \dfrac{\partial^2 h}{\partial x\,\partial y}(a) & \dfrac{\partial^2 h}{\partial y^2}(a) \end{bmatrix}.$$

The Hessian is a symmetric bilinear form on the tangent space $\mathrm{T}\mathbb{M}_a$ of $\mathbb{M}$ at a. The matrix above expresses this functional in terms of the basis $((\partial/\partial x)(a), (\partial/\partial y)(a))$ for $\mathrm{T}\mathbb{M}_a$. A critical point a is called *non-degenerate* if the Hessian is non-singular at a, i.e., $\det H(a) \neq 0$, a property that is independent of the coordinate system. The Morse Lemma [11] states that near a non-degenerate critical point a it is possible to choose local coordinates so that h takes the form

$$h(x, y) = h(a) \pm x^2 \pm y^2.$$

The number of minuses is called the *index* $i(a)$ of h at a; it equals the number of negative eigenvalues of $H(a)$ or, equivalently, the index of the functional $H(a)$. Note that the existence of these local coordinates implies that non-degenerate critical points are isolated.

In two dimensions there are three types of non-degenerate critical points: *minima* have index 0, *saddles* have index 1, and *maxima* have index 2. The function h is called a *Morse function* if all its critical points are non-degenerate. (Sometimes one also requires that the critical *values* of h, that is, the values that h takes at its critical points, are distinct. We will not need this requirement here.) Any twice differentiable function h can be unfolded to a Morse function.

Stable and Unstable Manifolds. In order to measure angles and lengths for tangent vectors, we choose a Riemannian metric $\langle , \rangle$ on $\mathbb{M}$, i.e., an inner product in each tangent space $\mathrm{T}\mathbb{M}_a$ that varies smoothly over $\mathbb{M}$. Since each vector in $\mathrm{T}\mathbb{M}_a$ is the tangent to a

curve γ in $\mathbb{M}$ through a, the *gradient* of h, ∇h, can be defined by the formula

$$\left\langle \frac{\mathrm{d}\gamma}{\mathrm{d}t}, \nabla h \right\rangle = \frac{\mathrm{d}(h \circ \gamma)}{\mathrm{d}t},$$

for every γ. It is always possible to choose coordinates (x, y) so that the tangent vectors $(\partial/\partial x)(a)$, $(\partial/\partial y)(a)$ are orthonormal with respect to $\langle , \rangle$. For such coordinates, the gradient is given by the familiar formula $\nabla h = ((\partial h/\partial x)(a), (\partial h/\partial y)(a))$.

An *integral line* $p\colon \mathbb{R} \to \mathbb{M}$ is a maximal path whose tangent vectors agree with the gradient, that is, $(\mathrm{d}/\mathrm{d}s)p(s) = \nabla h(p(s))$ for all $s \in \mathbb{R}$. The image of p is denoted by $\operatorname{im} p$. Each integral line is open at both ends. We call $\operatorname{org} p = \lim_{s\to-\infty} p(s)$ the *origin* and $\operatorname{dest} p = \lim_{s\to+\infty} p(s)$ the *destination* of the path p. These limits both exist because $\mathbb{M}$ is compact. Integral lines have the following three properties:

(P1) Two integral lines either have disjoint images or they are the same.
(P2) The images of integral lines cover all the non-critical points of $\mathbb{M}$.
(P3) The limits $\operatorname{org} p$ and $\operatorname{dest} p$ are critical points of h.

We use these properties to decompose $\mathbb{M}$ into regions of similar flow. The *stable manifold* $S(a)$ and the *unstable manifold* $U(a)$ of a critical point a are defined as

$$S(a) = \{a\} \cup \{y \in \mathbb{M} \mid y \in \operatorname{im} p, \operatorname{dest} p = a\},$$
$$U(a) = \{a\} \cup \{y \in \mathbb{M} \mid y \in \operatorname{im} p, \operatorname{org} p = a\}.$$

Note that the unstable manifolds of h are the stable manifolds of $-h$ as $\nabla(-h) = -\nabla(h)$. Therefore, the two types of manifolds have the same structural properties. An *open cell* of dimension i is a space homeomorphic to $\mathbb{R}^i$. The stable manifold $S(a)$ of a critical point a with index $i = i(a)$ is an open cell of dimension $\dim S(a) = i$. The closure of a stable manifold, however, is not necessarily homeomorphic to a closed ball, as seen in Fig. 1. By properties (P1)–(P3), the stable manifolds are pairwise disjoint and decompose $\mathbb{M}$ into open cells. The cells form a complex, as the boundary of every cell $S(a)$ is the union of lower-dimensional cells, its *faces*. The unstable manifolds similarly decompose $\mathbb{M}$ into a complex dual to the complex of stable manifolds: for $a, b \in \mathbb{M}$, $\dim S(a) = 2 - \dim U(a)$, and $S(a)$ is a face of $S(b)$ iff $U(b)$ is a face of $U(a)$.

Morse–Smale Complex. A Morse function h is a *Morse–Smale function* if the stable and unstable manifolds intersect only transversally. In two dimensions this means that stable and unstable 1-manifolds cross when they intersect. Their crossing point is necessarily a saddle, since crossing at a regular point would contradict property (P1). We intersect the stable and unstable manifolds to obtain the *Morse–Smale cells* as the connected components of the sets $U(a) \cap S(b)$, for all critical points $a, b \in \mathbb{M}$. The *Morse–Smale complex* is the collection of Morse–Smale cells. Note that $U(a) \cap S(a) = \{a\}$, and if $a \neq b$, then $U(a) \cap S(b)$ is the set of regular points $y \in \mathbb{M}$ that lie on integral lines p with $\operatorname{org} p = a$ and $\operatorname{dest} p = b$. It is possible that the intersection consists of more than one component, as seen in Fig. 1. We refer to the cells of dimension 0, 1, and 2 as *vertices*, *arcs*, and *regions*, respectively. Each vertex is a critical point, each arc is half of a stable or unstable 1-manifold, and each region is a component of the intersection of a stable and an unstable 2-manifold. We prove that the regions have a special shape.

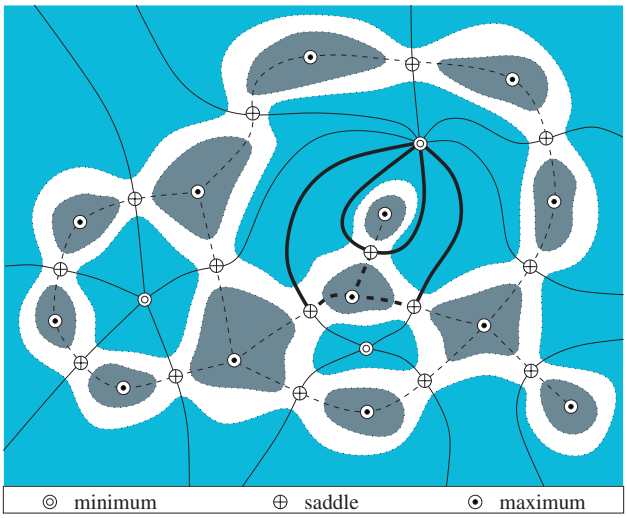

Fig. 1. A Morse–Smale complex with solid stable 1-manifolds and dashed unstable 1-manifolds. In drawing the dotted iso-lines we assume that all saddles have height between all minima and all maxima.

Quadrangle Lemma. *Each region of the Morse–Smale complex is a quadrangle with vertices of index* 0, 1, 2, 1, *in this order around the region. The boundary is possibly glued to itself along vertices and arcs.*

Proof. The vertices on the boundary of any region alternate between saddles and other critical points, which, in turn, alternate between maxima and minima. The shortest possible cyclic sequence of vertices around a boundary is therefore 0, 1, 2, 1, a quadrangle. We prove below that longer sequences force a critical point in the interior of the region, a contradiction.

We take a region whose boundary cycle has length $4k$ for $k \geq 2$ and glue two copies of the region together along their boundary to form a sphere. We glue each critical point to its copy, so saddles become regular points, maxima and minima remain as before. The Euler characteristic of the sphere is 2, and so is the alternating sum of critical points, $\sum_a (-1)^{i(a)}$. However, the number of minima and maxima together is $2k > 2$, which implies that there is at least one saddle inside the region. □

Quasi MS-Complexes. Intuitively, a quasi MS-complex is a complex with the structural form of a Morse–Smale complex. It is combinatorially a quadrangulation with vertices at the critical points of h and with arcs that strictly ascend or descend as measured by h. It differs from a Morse–Smale complex in that its arcs may not necessarily be the arcs of maximal ascent and descent. A subset of the vertices in a complex Q is *independent* if no two are connected by an arc. The complex Q is *splitable* if we can partition the vertices into three sets U, V, W and the arcs into two sets A, B so that

(i) $U \cup W$ and V are both independent,
(ii) arcs in A have endpoints in $U \cup V$ and arcs in B have endpoints in $V \cup W$, and

(iii) each vertex $v \in V$ belongs to four arcs, which in a cyclic order around v alternate between A and B.

We may then *split* Q into two complexes defined by U, A and W, B. Note that the Morse–Smale complex is splitable: (i) U, V, and W are the maxima, saddles, and minima, respectively, (ii) A connects maxima to saddles and B connects minima to saddles, and (iii) saddles have degree four and alternate as required. The Morse–Smale complex then splits into the complex of stable and the complex of unstable manifolds.

A *splitable quadrangulation* is a splitable complex whose regions are quadrangles. We define a *quasi MS-complex* of a 2-manifold $\mathbb{M}$ and a height function h as a splitable quadrangulation whose vertices are the critical points of h and whose arcs are monotonic in h.

3. PL 2-Manifolds

The gradient of a PL height function is not continuous and does not generate the pairwise disjoint integral lines that are needed to define stable and unstable manifolds. In this section we deal with the resulting difficulties by simulating differentiability using infinitesimal bump functions. Such a simulation unfolds degenerate critical points and turns stable and unstable manifolds into open cells, as needed.

Triangulation and Stars. Let K be a triangulation of a compact 2-manifold without boundary $\mathbb{M}$, and let $h\colon \mathbb{M} \to \mathbb{R}$ be a PL height function that is linear on every triangle. The height function is thus defined by its values at the vertices of K. It will be convenient to assume $h(u) \neq h(v)$ for all vertices $u \neq v$ in K. We simulate simplicity to justify this assumption computationally [8].

In a triangulation, the natural concept of a neighborhood of a vertex u is the *star*, $\operatorname{St} u$, that consists of u together with the edges and triangles that share u as a vertex. Formally, $\operatorname{St} u = \{\sigma \in K \mid u \leq \sigma\}$, where $u \leq \sigma$ is short for u being a face of σ. Since all vertices have different heights, each edge and triangle has a unique lowest and a unique highest vertex. Following Banchoff [2], we use this to define the *lower* and *upper stars* of u,

$$\underline{\operatorname{St}} u = \{\sigma \in \operatorname{St} u \mid h(v) \leq h(u), v \leq \sigma\},$$
$$\overline{\operatorname{St}} u = \{\sigma \in \operatorname{St} u \mid h(v) \geq h(u), v \leq \sigma\}.$$

These subsets of the star contain the simplices that have u as their highest or their lowest vertex. We may partition K into a collection of either subsets, $K = \dot{\bigcup}_u \underline{\operatorname{St}} u = \dot{\bigcup}_u \overline{\operatorname{St}} u$.

We may also use the lower and upper stars to classify a vertex as regular or critical. We define a *wedge* as a contiguous section of $\operatorname{St} u$ that begins and ends with an edge. As shown in Fig. 2, the lower star either contains the entire star or some number $k+1$ of wedges, and the same is true for the upper star. If $\underline{\operatorname{St}} u = \operatorname{St} u$, then $k = -1$ and u is a maximum. Symmetrically, if $\overline{\operatorname{St}} u = \operatorname{St} u$, then $k = -1$ and u is a minimum. Otherwise, u is regular if $k = 0$, a (simple) saddle if $k = 1$, and a *k-fold* or *multiple saddle* if $k \geq 2$. A two-fold saddle is often called a *monkey saddle*.

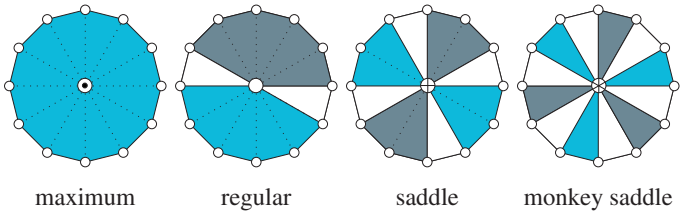

Fig. 2. The light shaded lower wedges are connected by white triangles to the dark shaded upper wedges.

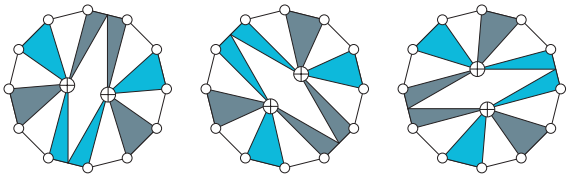

Fig. 3. A monkey saddle may be unfolded into two simple saddles in three different ways. If a wedge consists of a single edge, this edge unfolds into two copies.

Multiple Saddles. We can unfold a k-fold saddle into two saddles of multiplicity $1 \leq i, j < k$ with $i + j = k$ by the following procedure. We split a wedge of $\underline{\mathrm{St}}\, u$ (through a triangle, if necessary), and similarly split a non-adjacent wedge of $\overline{\mathrm{St}}\, u$. The new number of (lower and upper) wedges is $2(k + 1) + 2 = 2(i + 1) + 2(j + 1)$, as required. By repeating this process, we eventually arrive at k simple saddles. We place these saddles at the same height as the k-saddle they represent, and simulate perturbation. The combinatorial process is ambiguous, but for our purposes it is sufficient to pick an arbitrary unfolding from the set of possibilities. For a monkey saddle, there are three ways to unfold minimally, as shown in Fig. 3.

Merging and Forking. The concept of an integral line for a PL function is not well defined. Instead, we construct monotonic curves that never cross. Such curves can merge together and fork after a while. Moreover, it is possible for two curves to alternate between merging and forking an arbitrary number of times. To resolve this, when two curves merge, we pretend that they maintain an infinitesimal separation, running side by side without crossing. Figure 4 illustrates the two PL artifacts and the corresponding simulated smooth resolution. The computational simulation of disjoint integral lines is delicate and is described in Section 4.

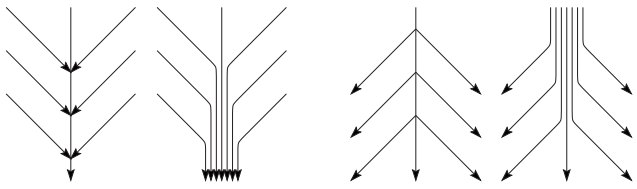

Fig. 4. Merging and forking PL curves and their corresponding smooth flow pictures.

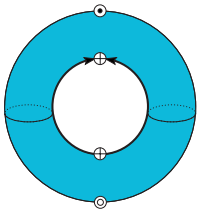

Fig. 5. The unstable 1-manifold of the lower saddle approaches the upper saddle.

Non-Transversal Intersections. The standard example in Morse theory is the height function over a torus standing on its side. The lowest and highest points of the inner ring are the only saddles, as shown in Fig. 5. Both the unstable 1-manifold of the lower saddle and the stable 1-manifold of the upper saddle follow the inner ring, so they overlap in two open half-circles. Generically, such non-transversal intersections do not happen. The characteristic property of a non-transversal intersection is that the unstable 1-manifold of one saddle approaches another saddle, and vice versa. An arbitrarily small perturbation of the height function suffices to make the two 1-manifolds miss the other saddles and approach a maximum and a minimum without meeting each other. The PL counterpart of a non-transversal intersection is an ascending or descending path that ends at a saddle. We simulate the generic case by extending the path beyond the saddle. Again, we give the details in Section 4.

4. Computing Quasi MS-Complexes

Given a triangulation K of a compact 2-manifold without boundary, and a PL height function h, our goal is to compute the Morse–Smale complex for a simulated unfolding of h. In this section we take a first step, computing a quasi MS-complex Q of h. To obtain a fast algorithm, we limit ourselves to paths using the edges of K. While the resulting complex is numerically inaccurate, the focus is on capturing the structure of the Morse–Smale complex.

Recall that the quasi MS-complex Q will have the critical points of h as vertices, and monotonic non-crossing paths as arcs. To resolve the merging and forking of paths, we formulate a three-stage algorithm. In each stage we compute a complex whose arcs are non-crossing monotonic paths, guaranteeing this property for the final complex.

Complex with Junctions. In the first stage we build a complex with extra vertices. We begin by classifying all vertices and computing the wedges of their lower and upper stars. We determine the steepest edge in each wedge and start $k+1$ ascending and $k+1$ descending paths from every k-fold saddle. Each path begins in its own wedge and follows a sequence of steepest edges until it hits

(a) a minimum or a maximum,
(b) a previously traced path at a regular point, or
(c) another saddle,

at which point the path ends. Case (a) corresponds to the generic case for smooth height functions, case (b) corresponds to a merging or forking, and case (c) is the PL counterpart of a non-transversal intersection between a stable and an unstable 1-manifold.

To resolve case (b), we allow regular points or *junctions* as vertices of the complex. We either create a new junction and split the previously traced path, or we increase the degree of the previously created junction. By definition, junctions remove all crossings in the complex. We will eliminate junctions and resolve case (c) in the second stage of the algorithm.

We use the quad edge data structure [10] to store the complex defined by the paths. The vertices of the complex are the critical points and junctions, and the arcs are the pairwise edge-disjoint paths connecting these vertices.

Extending Paths. In the second stage of our algorithm, we extend paths to remove junctions and reduce the number of arcs per k-fold saddle to $2(k+1)$. Whenever we extend a path, we route it along and infinitesimally close to an already existing path. In practice, we simulate this extension combinatorially within the quad-edge data structure. In extending paths, we may create new paths ending at other junctions and saddles. Consequently, we must process the vertices in a sequence that prevents cyclic dependencies. Since ascending and descending paths are extended in opposite directions, we need two orderings and we touch every vertex twice. It is convenient first to extend ascending paths in the order of increasing height, and second to extend descending paths in the order of decreasing height. We next discuss our routing procedures for junctions and saddles. In the figures that follow, we orient paths in the direction they emanate from a saddle.

Consider the junction y in Fig. 6 on the left. By definition, y is a regular point with lower and upper stars consisting of one wedge each. The first time we encounter y, the path is traced right through the point. In every additional encounter, the path ends at y, as y is now a junction. If the first path is ascending, then one ascending path leaves y into the upper star, all other ascending paths approach y from the lower star, and all descending paths approach y from the upper star. This is the case shown in Fig. 6. We duplicate paths for all junctions using our two orderings. Note that the new paths, shown in the middle of Fig. 6, may include duplicates spawned by junctions that occur before this vertex in an ordering. Finally, we concatenate the resulting paths in pairs without creating crossings, as shown in Fig. 6 to the right.

We next resolve case (c), paths that have another saddle as an endpoint. Consider the saddle x in Fig. 7. We look at path extensions only within one of the sectors between two

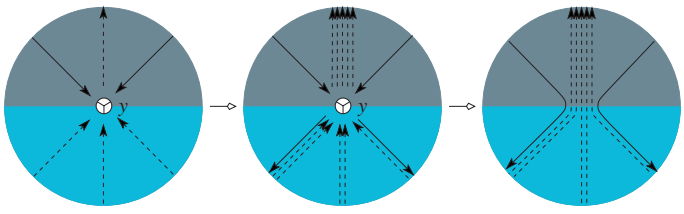

Fig. 6. Paths ending at junctions are extended by duplication and concatenation.

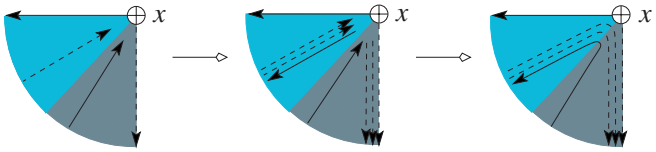

Fig. 7. Paths that end at a saddle by case (c) are extended by duplication and concatenation.

cyclically contiguous steepest edges. Within this sector there may be ascending paths approaching x from within the overlapping wedge of the lower star, and descending paths approaching x from within the overlapping wedge of the upper star, as shown in Fig. 7 to the left. After path duplications, we concatenate the paths in pairs. Again, we can concatenate without creating crossings. At the end of this process, our complex has critical points as vertices, and monotonic non-crossing paths from saddles to minima or maxima as arcs.

Unfolding Multiple Saddles. In the third and last stage of our algorithm, we unfold every k-fold saddle into k simple saddles. As indicated in Fig. 3, we can do so simply by duplicating the saddle and paths ending at the saddle. All paths of case (c) have already been removed in the second stage, so we only have to deal with the $k+1$ ascending and $k+1$ descending paths that originate at the k-fold saddle. In each of the $k-1$ steps, we duplicate the saddle, one ascending path, and a non-adjacent descending path. In the end, we have k saddles and $2(k+1)+2(k-1) = 4k$ paths, or four per saddle. Figure 8 illustrates the operation by showing a possible unfolding of a three-fold saddle. The unfolding procedure does not create any path crossings in the previous complex, which had no crossings.

Quasi MS-Complex Lemma. *The algorithm computes a quasi MS-complex for K.*

Proof. Let Q be the complex constructed by the algorithm. The vertices of Q are the unfolded critical points of K, so they are minima, saddles, and maxima. The paths are non-crossing and stage two guarantees that the paths go from saddles to minima or maxima. Therefore, Q is splitable. Moreover, the vertices on the boundary of any region of Q alternate between saddles and other critical points. The Quadrangle Lemma implies Q is a quadrangulation. Therefore, Q is a splitable quadrangulation, or a quasi MS-complex. □

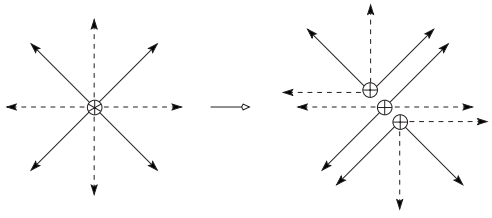

Fig. 8. Unfolding a three-fold saddle into three simple saddles.

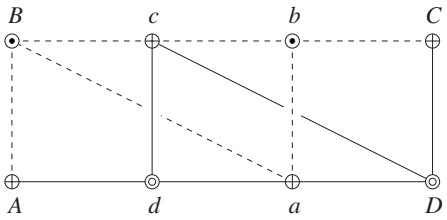

Fig. 9. The octagon is the union of a row of three quadrangles. The haloed edges indicate the alternative quadrangulation created by the handle slide.

5. Local Transformation

We transform a quasi MS-complex to the Morse–Smale complex via a sequence of transformations called handle slides. We first describe these transformations, and then present and analyze an algorithm that applies handle slides to the quasi MS-complex.

Handle Slide. A handle slide transforms one quasi MS-complex into another. The two quadrangulations differ only in their decompositions of a single octagon. In the first quadrangulation the octagon consists of a quadrangle *abcd* together with two adjacent quadrangles *baDC* and *dcBA*. Assume that *d* and *D* are minima and *b* and *B* are maxima, as in Fig. 9. We perform a slide by drawing an ascending path from *a* to *B* replacing *ab*, and a descending path from *c* to *D* replacing *cd*. After the slide, the octagon is decomposed into quadrangles *DcBa* in the middle and *cDCb*, *aBAd* on its two sides.

It is possible to think of the better known edge-flip in a two-dimensional triangulation as the composition of two octagon slides. To explain this, we superimpose a triangulation with its dual diagram, making sure that only corresponding edges cross, as in Fig. 10. The vertices of the triangulation correspond to minima, the vertices of the dual diagram to maxima, and the crossing points to saddles. When we flip an edge in the triangulation, we also reconnect the five edges in the dual diagram that correspond to the five edges of the two triangles sharing the flipped edge. The result of the edge-flip is thus the same as that of two octagon slides, one for the lower left three quadrangles, and the other for the upper right three quadrangles in Fig. 10.

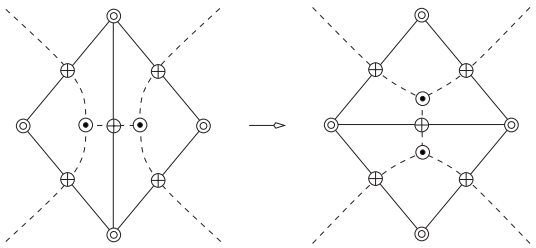

Fig. 10. Edge-flip shown in super-imposition of solid triangulation with its dashed dual diagram. The maxima before and after the flip should be at the same location but are moved for clarity of the illustration.

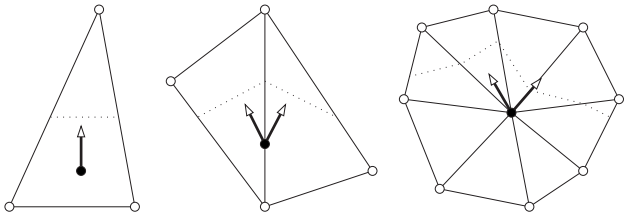

Fig. 11. The directions of locally steepest ascent are orthogonal to the dotted level lines.

Steepest Ascent. We decide whether or not to apply a handle slide to an octagon by rerouting its interior paths. We reroute an ascending path by following the direction of locally steepest ascent, which may go along an edge or pass through a triangle of K. There are three cases as shown in Fig. 11. In the interior of a triangle uvw, that steepest direction is unique and orthogonal to the level lines. In the interior of an edge, there may be one or two locally steepest directions, and at a vertex there may be as many locally steepest directions as there are triangles in the star. We may compute the globally steepest direction numerically with small error, but errors accumulate as the path traverses triangles. Alternatively, we can compute the globally steepest direction exactly with constant bit-length arithmetic operations, but the bit-length needed for the points along the path grows as it traverses more triangles. This phenomenon justifies the SoD approach to constructing a Morse–Smale complex. In that approach the computed complex has the same combinatorial form as the Morse–Smale complex, and it is numerically as accurate as the local rerouting operations used to control handle slides.

Algorithm. We now describe how transformations are applied to construct the Morse–Smale complex. The algorithm applies handle slides in the order of decreasing height, where the *height* of an octagon is the height of the lower saddle of the middle quadrangle. In Fig. 9 this saddle is either a or c, and we assume here that it is a. At the time we consider a, we may assume that the arcs connecting higher critical points are already correct. The iso-line at the height of a decomposes the manifold into an *upper* and a *lower region*, and we let Γ be the possibly pinched component of the upper region that contains a. There are two cases. In the first case, which is illustrated in Fig. 12, the higher critical points in Γ and their connecting arcs bound one annulus, which is pinched at a. In the second case, which is illustrated in Fig. 13, these arcs bound two annuli, one on each side of a. The ascending paths emanating from a are rerouted within these annuli.

Let ab be the interior path of the octagon with height $h(a)$, and let p be the maximum we hit by rerouting the path. If p is the first maximum after b along the arc boundary of the annulus, we may use a single handle slide to replace ab by ap, as for ap_1 in Fig. 12. Note that the slide is possible only because ap_1 crosses no arc ending at b_1. Any such arc would have to be changed first, which we do by recursive application of the algorithm, as for ap_2 in Fig. 12. It is also possible that p is more than one position removed from b, as for ap_1 in Fig. 13. In this case we perform several slides for a, the first connecting a to the first maximum after b in the direction of p. Each such slide may require recursive slides to clear the way, as before. Finally, it is possible that the new path from a to p

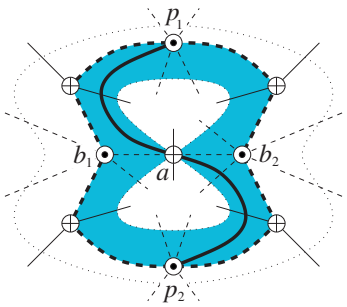

Fig. 12. The case of a single annulus pinched at a. The iso-line is dotted, the annulus is shaded, the arcs bounding the annulus are bold dashed, and the new paths emanating from a are bold solid.

winds around the arc boundary of the annulus several times, as does ap_2 in Fig. 13. The algorithm is the same as before.

The winding case shows that the number of slides cannot be bounded from above in terms of the number of critical points. Instead we consider crossings between arcs of the initial quasi MS-complex and the final Morse–Smale complex, and note that the number of slides is at most some constant times the number of such crossings.

6. Topological Persistence

We may measure the importance of every critical point and construct a hierarchy of complexes by eliminating critical points with measure below a threshold. In this section we describe a measure called persistence and discuss its computation. Further details may be found in [7].

Filtration and Betti Numbers. Assume we assemble the triangulation K of a connected compact 2-manifold without boundary $\mathbb{M}$ by adding simplices in the order of increasing height. Let $u^1, u^2, \ldots, u^n$ be the sequence of vertices such that $h(u^i) < h(u^j)$ for all $1 \leq i < j \leq n$, and let K^j be the union of the first j lower stars, $K^j = \bigcup_{1 \leq i \leq j} \underline{\mathrm{St}}\, u^i$. The subcomplex K^j of K consists of the j lowest vertices together with all edges and triangles connecting them. We call $K^1, K^2, \ldots, K^n$ a *filtration* of K. Note that this

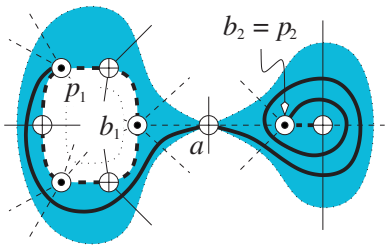

Fig. 13. The case of two annuli connected at a. We use dotted, dashed, and solid lines as in Figure 12.

definition of filtration comes from combinatorial topology and is different from the one used in dyamical systems [15]. Let β_0^j, β_1^j, and β_2^j be the three possibly non-trivial Betti numbers of K^j. Assuming an orientable 2-manifold $\mathbb{M}$, $\beta_0 = \beta_0^n = 1$, $\beta_1 = \beta_1^n$, and $\beta_2 = \beta_2^n = 1$ are the Betti numbers of $K = K^n$. The Betti numbers of K^{j+1} can be computed from those of K^j merely by looking at the type of u^{j+1} and how its lower star connects to K^j [6]. It is convenient to adopt reduced homology groups, but we will freely talk about components and holes when we mean reduced homology classes of non-bounding 0- and 1-cycles. We start with $\beta_{-1}^0 = 1$ and $\beta_0^0 = \beta_1^0 = \beta_2^0 = 0$. Regular vertices u^{j+1} can be skipped because they do not change the Betti numbers.

Case 0: u^{j+1} *is a minimum*. If $j + 1 = 1$, then $\beta_{-1}^1 = \beta_{-1}^0 - 1 = 0$. Otherwise u^{j+1} forms a new component and we increment the zeroth Betti number to $\beta_0^{j+1} = \beta_0^j + 1$.

Case 1: u^{j+1} *is a k-fold saddle*. The lower star touches K^j along $k+1$ simple paths. Let $1 \leq \gamma \leq k + 1$ be the number of touched components. Then the lower star decreases the number of components to $\beta_0^{j+1} = \beta_0^j - (\gamma - 1)$ and increases the number of holes to $\beta_1^{j+1} = \beta_1^j + (k + 1 - \gamma)$.

Case 2: u^{j+1} *is a maximum*. If $j + 1 = n$, then adding the lower star completes the manifold and $\beta_2^n = \beta_2^{n-1} + 1 = 1$. Otherwise, the lower star closes a hole and we decrement the first Betti number to $\beta_1^{j+1} = \beta_1^j - 1$.

The Betti numbers change as components and holes are created and destroyed. Every minimum, except the first, creates a non-bounding 0-cycle. Every maximum, except the last, destroys a non-bounding 1-cycle. Every simple saddle either destroys a component, if $\gamma = 2$, or creates a hole, if $\gamma = 1$. A k-fold saddle has the accumulated effect of the k simple saddles.

Persistence. The idea of persistence is the realization that acts of creation can be paired with acts of destruction. Call a critical point *positive* if it creates and *negative* if it destroys. Assume that all saddles are simple or, equivalently, that all multiple saddles have been unfolded. We pair every negative saddle with a preceding positive minimum and every negative maximum with a preceding positive saddle.

To determine the pairs, we scan the filtration from left to right. A positive minimum starts and represents a new component. A negative saddle connects two components of the complex it is added to. Each component is represented by its lowest minimum, and the saddle is paired with the higher of the minima. The other yet unpaired minimum lives on as the representative of the merged component. A positive saddle starts and represents a new non-bounding cycle. A negative maximum fills a hole in the complex. The boundary of that hole is homologous to a sum of cycles, each represented by a positive saddle. The maximum is paired with the highest of these saddles, and the other saddles live on as representatives of their respective cycles.

At the end of this process, we have minimum-saddle pairs, saddle-maximum pairs, and a collection of $\beta_0 = 1$ minima, β_1 simple saddles, and $\beta_2 = 1$ maxima that remain unpaired. The *persistence* of a critical point a is the absolute height difference $p(a) = |h(b) - h(a)|$, if a is paired with b, and $p(a) = \infty$ if a remains unpaired. We illustrate the pairing and the resulting persistence by drawing the critical points on the (horizontal) height axis as shown in Fig. 14.

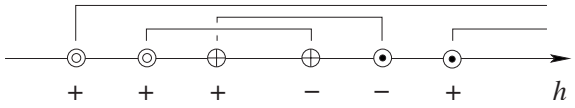

Fig. 14. Each critical point is either positive or negative. The persistence is the absolute height difference between paired critical points.

We may also assemble a triangulation in the order of decreasing height. Define L^{n-j+1} as the union of the upper stars of the j highest vertices. The sequence $L^n, L^{n-1}, \ldots, L^1$ is again a filtration of K, and we can compute Betti numbers as before. Because of the reversal of direction, minima and maxima exchange their roles in the Betti number algorithm, and negative critical points act like positive ones and vice versa. Although everything is reversed, the persistence of critical points remains unchanged. In other words, we have the same pairing of critical points, regardless of the direction of assembly for the filtration.

Computation. We compute the pairing of critical points and their persistence using the algorithm given in [7]. Instead of applying that algorithm to a sequence of critical points (as described above), we apply it to a sequence of simplices. In the ascending direction, the appropriate sequence is obtained by replacing each critical point u^i by the simplices in its lower star, ordered by non-decreasing dimension. The persistence algorithm pairs most simplices in a lower star with other simplices in that set. The number and dimensions of the simplices that are not paired within the lower star are characteristic for the type of the critical point u^i. Specifically, there is one unpaired vertex if u^i is a minimum, there are k unpaired edges if u^i is a k-fold saddle, and there is one unpaired triangle if u^i is a maximum. We use this characterization to find and classify critical points in practice.

The algorithm that pairs critical points is similar to but different from the constructive definition given above. The minimum-saddle pairs are found by scanning the filtration from left to right. For each negative saddle we determine the matching positive minimum by searching backwards. With the help of a union-find data structure storing the components, this can be done in time $O(nA^{-1}(n))$, where $A^{-1}(n)$ is the notoriously slow growing inverse of the Ackermann function. Symmetrically, the saddle-maximum pairs are found by scanning the reversed sequence of simplices that corresponds to the filtration of the L^j. The running time is again $O(nA^{-1}(n))$. Note that the two-scan algorithm makes essential use of the fact that persistence is symmetric with respect to increasing and decreasing height. We could also find the saddle-maximum pairs in the first scan, but the running time would be worse as the search for the matching saddle in that direction cannot be accelerated by using a union-find data structure.

7. Hierarchy

Given a Morse–Smale complex, we create a hierarchy by successive simplification. Each step in the process cancels a pair of critical points and the sequence of cancellations is determined by the persistence of the pairs.

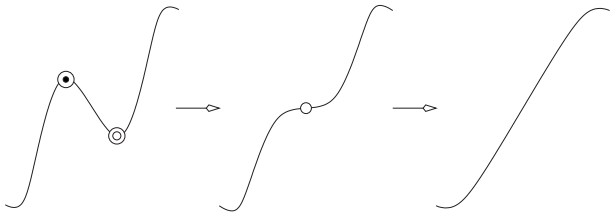

Fig. 15. The cancellation of a minimum–maximum pair.

Cancellation. To simplify the discussion, consider first a generic one-dimensional height function $f\colon \mathbb{R} \to \mathbb{R}$. Its critical points are minima and maxima in an alternating sequence from left to right. In order to eliminate a maximum, we locally modify f so that the maximum moves towards an adjacent minimum. When the two points meet, they momentarily form a degenerate critical point and then disappear, as illustrated in Fig. 15. Clearly, only adjacent critical points can be canceled, but adjacency is not sufficient unless we are willing to modify f globally. We also require that the height difference between the two critical points is less than that of pairs in the neighborhood. In Fig. 16 the pairs computed by the persistence algorithm and plotted along the domain axis are either disjoint or nested. We cancel pairs of critical points in the order of increasing persistence. The nesting structure is thus unraveled from inside out by removing one innermost pair after the other.

Simplification. We now return to our height function h over $\mathbb{M}$. The critical points of h can be eliminated in a very similar manner by locally modifying the height function. In the generic case, the critical points cancel in pairs of contiguous indices. More precisely, positive minima cancel with negative saddles and positive saddles cancel with negative maxima. We simulate the cancellation process combinatorially by removing critical points in pairs from the Morse–Smale complex. Figure 17 illustrates the operation for a minimum b paired with a saddle a. The operation requires that ab be an arc in the complex. Let c be the other minimum and let d, e be the two maxima connected to a. We delete the two ascending paths from a to d and e, and contract the two descending paths from a to b and c. In the symmetric case in which b is a maximum, we delete

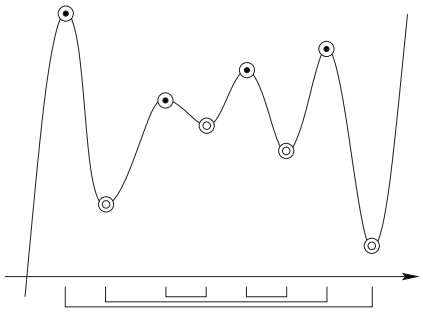

Fig. 16. The intervals defined by critical point pairs are either disjoint or nested.

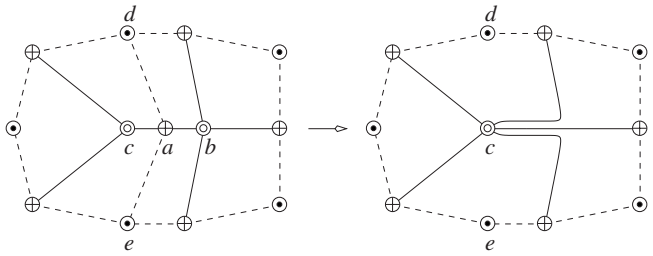

Fig. 17. The cancellation of a and b deletes the arcs ad and ae and contracts the arcs ca and ab. The contraction effectively extends the remaining arcs of b to c.

the descending and contract the ascending paths. The contraction pulls a and b into the critical point c, which inherits the connections of b. We call this operation the *cancellation* of a and b. It is the only operation needed in the construction of the hierarchy. There are two special cases, namely, when $d = e$ and when $b = c$. In the latter case we prohibit the cancellation as it would change the topology of the 2-manifold.

The sequence of cancellations is again in the order of increasing persistence. In general, not every two critical points paired by the persistence algorithm are adjacent in the Morse–Smale complex. We prove below that they will be adjacent at the required time.

Adjacency Lemma. *For every positive i, the i*th *pair of critical points ordered by persistence forms an arc in the complex obtained by canceling the first $i - 1$ pairs.*

Proof. Assume without loss of generality that the ith pair consists of a negative saddle $a = u^{j+1}$ and a positive minimum z. Consider the component of K^j that contains z. One of the descending paths originating at a enters this component and because it cannot ascend it eventually ends at some minimum b in the same component. Either $b = z$, in which case we are done, or b has already been paired with a saddle $c \neq a$. In the latter case, c has height less than a, it belongs to the same component of K^j as b and z, and b, c is one of the first $i - 1$ pairs of critical points. It follows that when b gets canceled, the path from a to b gets extended to another minimum d, which again belongs to the same component. Eventually, all minima in the component other than z are canceled, implying that the initial path from a to b gets extended all the way to z. The claim follows. □

Note that the proof also works for quasi MS-complexes, which implies that the Adjacency Lemma is also valid for quasi MS-complexes.

8. Results for Terrains

This section presents experimental results to support the viability of our approach for analyzing geographic terrain data. At this time, we have only implemented the algorithms for constructing quasi MS-complexes and the persistence of critical points.

Table 1. The five data sets. The second column counts all vertices, edges, and triangles of the triangulations.

	Grid size	Number of simplices
Sine	100×100	59,996
Iran	277×229	380,594
Himalayas	469×265	745,706
Andes	385×877	2,025,866
North America	793×505	2,402,786

Data Sets. We use four rectangle sections of rectilinear grid elevation data of Earth [14] and one synthetic data sampled from $h(x, y) = \sin x + \sin y$ for input. The names and sizes of the data sets are given in Table 1. In each case we convert the gridded rectangle into a triangulated sphere by adding diagonals to the square cells and connecting the boundary edges and vertices to a dummy vertex at height minus infinity. We show the quasi MS-complex of data set Sine in Fig. 18. It is computed by the algorithm presented in Section 4, and it is also the Morse–Smale complex, in this case. In Fig. 19 we display the terrain of Iran along with its quasi MS-complex.

Statistics. We first compute a filtration of the sphere triangulation by sorting the simplices in the order of increasing height, as explained in Section 6. We then use the persistence algorithm to pair all simplices, identifying and classifying the critical points as a side-product.

Table 2 shows the number of critical points of each type. Since we start with grid data and add diagonals in a consistent manner, each vertex other than the dummy vertex has degree at most six. We can therefore have monkey saddles in our data but no saddles of

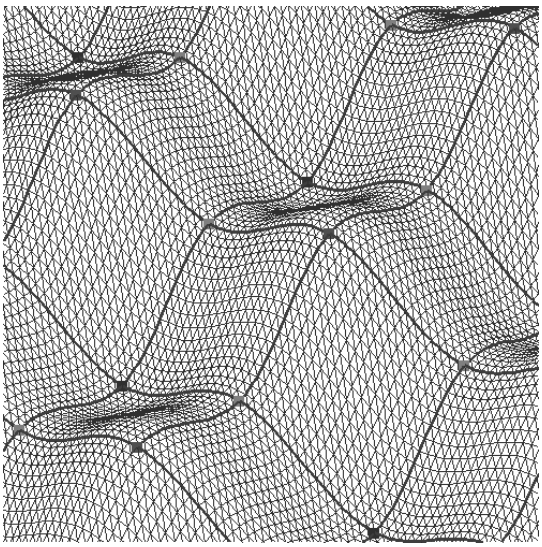

Fig. 18. The Morse–Smale complex partitions the triangulated data sampled from $h(x, y) = \sin x + \sin y$.

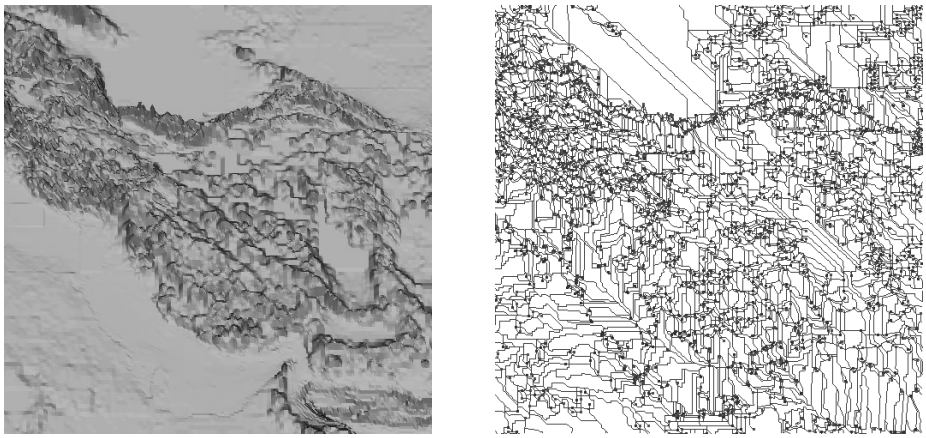

Fig. 19. Iran's Alburz mountain range borders the Caspian sea (top flat area), and its Zagros mountain range shapes the Persian Gulf (left bottom). We show a rendering of the terrain and its quasi MS-complex.

multiplicity higher than two. The current implementation constructs only one filtration and computes persistence in a single scan, forfeiting the benefit of a union-find data structure for pairing maxima with saddles. Table 3 shows the running time of constructing the filtration, computing the persistence information, and constructing the quasi MS-complex.

9. Conclusion

This paper introduces Simulation of Differentiability as a computational paradigm, and uses it to construct Morse–Smale complexes of PL height functions over compact 2-manifolds without boundary. It also uses topological persistence to build a hierarchy of progressively coarser Morse–Smale complexes.

Our results complement and improve related work in visualization [1], [5], [16] and computational geometry [3], [4], [17]. The terrain simplification procedure described exploits the relationship between the geometry of the terrain and the topology of its contours. As such it is different from simplification algorithms guided by purely geometric numerical measures, such as the quadrics metric [9]. Our algorithm preserves

Table 2. The number of critical points of the five triangulated spheres. (Note that #Min − #Sad − 2#Mon + #Max = 2 in each case, as it should be.)

	# Min	# Sad	# Mon	# Max
Sine	10	24	0	16
Iran	1,302	2,786	27	1,540
Himalayas	2,132	4,452	51	2,424
Andes	20,855	38,326	1,820	21,113
North America	15,032	30,733	464	16,631

Table 3. Running times in seconds. (All timings were done on a Sun Ultra-10 with a 440 MHz UltraSPARC IIi processor and 256 megabyte RAM, running the Solaris 8 operating system.)

	Filt.	Pers.	qMS
Sine	0.06	0.13	0.03
Iran	0.46	0.90	0.56
Himalayas	0.89	1.74	1.01
Andes	2.62	4.90	2.60
North America	3.28	5.84	5.26

important critical points of the terrain, making it appropriate for applications in geographic information systems (GIS), such as computing water flow routing and accumulation [13].

Many questions remain. Can we go from a quasi MS-complex to the Morse–Smale complex by performing handle slides in an arbitrary sequence, as opposed to ordering them by height? How can we take advantage of the Morse–Smale complex structure and cancel critical points in a controlled manner through smoothing or locally averaging the height function? What are the dependencies between different critical points, and how does a cancellation interfere with non-participating critical points?

There are at least three interesting challenges in generalizing the results of this paper. Can we use hierarchical methods similar to the ones in this paper to simplify medial axes of two-dimensional figures? Can we extend our methods to more general vector fields? Can we extend our results to height functions over 3-manifolds? The generalization of the Morse index of critical points to the Conley index of isolated neighborhoods [12] might be useful in answering the second question.

References

1. Bajaj, C. L., Pascucci, V., and Schikore, D. R. Visualization of scalar topology for structural enhancement. In *Proc. 9th Ann. IEEE Conf. Visualization* (1998), pp. 18–23.
2. Banchoff, T. F. Critical points and curvature for embedded polyhedral surfaces. *Amer. Math. Monthly* **77** (1970), 475–485.
3. Carr, H., Snoeyink, J., and Axen, U. Computing contour trees in all dimensions. In *Proc. 11th Ann. Sympos. Discrete Algorithms* (2000), pp. 918–926.
4. de Berg, M., and van Kreveld, M. Trekking in the alps without freezing and getting tired. In *Proc. 1st Europ. Sympos. Algorithms* (1993), pp. 121–132.
5. de Leeuw, W., and van Liere, R. Collapsing flow topology using area metrics. In *Proc. 10th Ann. IEEE Conf. Visualization* (1999), pp. 349–354.
6. Delfinado, C. J. A., and Edelsbrunner, H. An incremental algorithm for Betti numbers of simplicial complexes on the 3-sphere. *Comput. Aided Geom. Design* **12** (1995), 771–784.
7. Edelsbrunner, H., Letscher, D., and Zomorodian, A. Topological persistence and simplification. *Discrete Comput. Geom.* **28** (2002), 511–533.
8. Edelsbrunner, H., and Mücke, E. P. Simulation of Simplicity: a technique to cope with degenerate cases in geometric algorithms. *ACM Trans. Graphics* **9** (1990), 66–104.
9. Garland, M., and Heckbert, P. S. Surface simplification using quadric error metrics. In *SIGGRAPH 97 Conf. Proc.* (1997), pp. 209–216.

10. Guibas, L., and Stolfi, J. Primitives for the manipulation of general subdivisions and the computation of Voronoi diagrams. *ACM Trans. Graph.* **4** (1985), 74–123.
11. Milnor, J. *Morse Theory*. Princeton University Press, Princeton, NJ, 1963.
12. Mischaikow, K., and Mrozek, M. Conley index theory. In *Handbook of Dynamical Systems II*, B. Fiedler, Ed. Elsevier, Amsterdam, 2002.
13. Moore, I. D., Grayson, R. B., and Ladson, A. R. Digital terrain modeling: a review of hydrological, geomorphological, and biological applications. *Hydrological Process.* **5** (1991), 3–30.
14. NOAA. Announcement 88-mgg-02, digital relief of the surface of the earth, 1988. http://www.ngdc.noaa.gov/mgg/global/seltopo.html.
15. Palis, J., and de Melo, W. *Geometric Theory of Dynamical Systems*. Springer-Verlag, New York, 1982.
16. Tricoche, X., Scheuermann, G., and Hagen, H. A topology simplification method for 2D vector fields. In *Proc. 11th Ann. IEEE Conf. Visualization* (2000), pp. 359–366.
17. van Kreveld, M., van Oostrum, R., Bajaj, C., Pascucci, V., and Schikore, D. Contour trees and small seed sets for iso-surface traversal. In *Proc. 13th Ann. Sympos. Computational Geometry* (1997), pp. 212–220.
18. Wallace, A. *Differential Topology. First Steps*. Benjamin, New York, 1968.

Received June 14, 2001, *and in revised form September* 11, 2002. *Online publication May* 7, 2003.