LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

Topology-based Simplification for Feature Extraction from 3D Scalar Fields

A. Gyulassy, V. Natarajan, V. Pascucci, P.-T. Bremer, B. Hamann

October 15, 2005

Disclaimer

Topology-based Simplification for Feature Extraction from 3D Scalar Fields

Attila Gyulassy*   Vijay Natarajan†   Valerio Pascucci‡   Peer-Timo Bremer§   Bernd Hamann¶
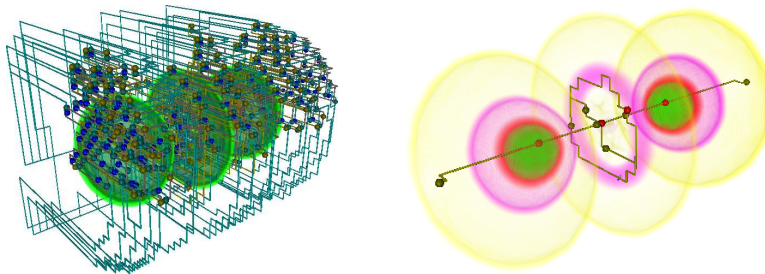
Figure 1: Topology simplification applied on electron density data for a hydrogen atom: the input has a large number of critical points, several of which are identified as being insignificant and removed by repeated application of two atomic operations. Features are identified by the surviving critical points and enhanced in a volume rendered image by an automatically designed transfer function.

ABSTRACT

This paper describes a topological approach for simplifying continuous functions defined on volumetric domains. We present a combinatorial algorithm that simplifies the Morse-Smale complex by repeated application of two atomic operations that removes pairs of critical points. The Morse-Smale complex is a topological data structure that provides a compact representation of gradient flows between critical points of a function. Critical points paired by the Morse-Smale complex identify topological features and their importance. The simplification procedure leaves important critical points untouched, and is therefore useful for extracting desirable features. We also present a visualization of the simplified topology.

CR Categories: I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling; I.3.6 [Computer Graphics]: Methodology and Techniques;

Keywords: Morse theory, Morse-Smale complexes, computational topology, multiresolution, simplification, feature detection, 3D scalar fields.

1 INTRODUCTION

Scientific data are becoming increasingly complex and require sophisticated hierarchical representations for their effective exploration. In the process of creating such a hierarchy, it is highly desirable that important features be identified and preserved while insignificant features be removed early. Note that noise in the data

---
*Institute for Data Analysis and Visualization, Dept. of Computer Science, University of California, Davis. e-mail: aggyulassy@ucdavis.edu

†Institute for Data Analysis and Visualization, Dept. of Computer Science, University of California, Davis. e-mail: vnatarajan@ucdavis.edu

‡Center for Applied Scientific Computing, Lawrence Livermore National Laboratory. e-mail: pascucci1@llnl.gov

§Dept. of Computer Science, University of Illinois, Urbana Champaign. e-mail: ptbremer@acm.org

¶Institute for Data Analysis and Visualization, Dept. of Computer Science, University of California, Davis. e-mail: hamann@cs.ucdavis.edu

could also be interpreted as a feature, albeit one with relatively low importance. A crucial ingredient of any hierarchy building process is the ordering of features based on significance. Current methods typically adopt a geometric approach where the numerical error associated with the simplified model is used as the measure of approximation quality. Any removal of topological features caused by these methods is incidental and not always controlled. We adopt a more direct approach by explicitly identifying the topological features, ordering them, and removing features in order of increasing significance.

Related work. Scientific data is usually represented as a set of discrete samples of a function defined on a two- or three-dimensional domain. A continuous function is obtained by interpolating the samples within an underlying mesh of the domain. In order to efficiently handle the increasingly large datasets for visualization purposes, several methods have been proposed to simplify them *i.e.*, reduce the domain size and complexity. These methods typically decimate the mesh by repeated application of a fundamental operation, the most successful one being edge contraction [13]. The error introduced by edge contraction is computed as the sum of distances to planes that are associated with end points of the contracted edge [10]. Originally developed for surface meshes, edge contraction has been successfully extended to tetrahedral meshes [4, 16] and higher-dimensional meshes [11]. A purely geometric approach to simplification is compatible with but cannot replace topology simplification. Naturally, there has been considerable work done on topological simplification of scalar functions. In the case of a real-valued function defined on a 3D domain, the topological features are created/destroyed by critical points of the function. Two data structures have been proposed in the literature for storing these topological features: Reeb graphs and Morse-Smale complexes. The *Reeb graph* [17] traces components of isocontours/isosurfaces as they sweep the domain. In the case of simply connected domains, the Reeb graph has no cycles and is called a *contour tree*. Reeb graphs, contour trees, and their variants have been used quite successfully to guide the removal of topological features [3, 12, 20, 22]. The *Morse-Smale complex* [19] decomposes the domain of a function into regions having uniform gradient flow behavior. It has been used recently to perform controlled simplification of topology for functions defined on 2-manifolds [2]. The use of Morse-Smale complex to guide data simplification allows this technique to utilize a global view of the function for detecting, ordering and re-
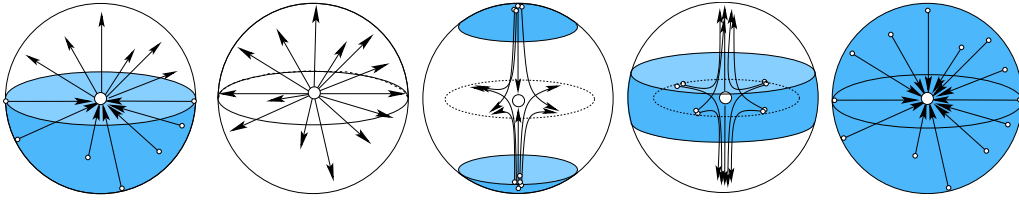
Figure 2: Local pictures of a regular point and the four types of critical points (minimum, 1-saddle, 2-saddle, and maximum) with shaded oceans, white continents and integral lines.

moving features along with the ability to perform the simplification within a local neighborhood of the non-significant feature. Reeb graph based simplification methods, on the other hand, work within a local neighborhood of the domain both during feature identification and simplification stages. Furthermore, in 3D, they are limited to detecting and simplifying features that are associated with the creation and destruction of isosurface components. These features are represented by a pair of critical points consisting of one saddle and one extremum. The Morse-Smale complex is able to detect genus changes within the isosurface as well, which are represented by a saddle-saddle pair. We take this more comprehensive approach for simplifying functions defined on 3-manifolds. Simplification of topological features has been studied in the context of vector fields also [6, 21]. These methods, however, use numerical approaches to identify, order, and simplify the topology and are therefore prone to instability. In contrast, Reeb graph and Morse-Smale complex based methods for simplification of scalar fields are combinatorial in nature and hence stable.

Results. We describe a new method for explicit removal of topological features in a given 3D scalar function with the goal of constructing a hierarchical representation. We extend the results of Bremer et al. [2] and Edelsbrunner et al. [8] to functions defined on 3-manifolds by following their approach and introducing new ideas to address issues that arise due to the added complexity of features in 3D domains. The extension is made non-trivial by the presence of two different types of saddles in 3-manifolds. For 2-manifolds, simplification of the function using the Morse-Smale complex is accomplished by repeated cancellation of critical point pairs, one of which is a saddle and the other an extremum (minimum/maximum). The presence of a new type of saddle in 3D creates the need for an additional type of cancellation. We introduce a saddle-saddle cancellation and extend the saddle-extremum cancellation to 3D, to obtain two atomic operations that simplify the function. Repeated application of these two operations on ordered pairs of critical points results in a hierarchy. Critical point pairs are ordered based on the notion of persistence [9], which measures the importance of the associated topological feature. We apply our method to data obtained from various application areas to demonstrate the power of our approach.

2 Morse Theory

Our algorithms are based on Morse theory, which was originally developed for smooth functions. In this section, we give a brief description of Morse theoretic ideas essential to this paper and their extension to piecewise-linear (PL) functions defined over tetrahedral meshes. We refer to the paper by Edelsbrunner et al. [7] for a comprehensive description of this extension to PL functions. Morse theory is discussed in detail by Milnor [15] and Matsumoto [14].

Critical points. Let $\mathbb{M}$ be a compact 3-manifold and $f : \mathbb{M} \to \mathbb{R}$ be a real-valued smooth map defined on $\mathbb{M}$. f is a *Morse function* if none of its critical points are degenerate (*i.e.*, the Hessian matrix is

non-singular for all critical points) and no two critical points have the same function value. The Morse Lemma states that a Morse function has quadratic behavior within a local neighborhood of every non-degenerate critical point p. This lemma immediately characterizes critical points. Figure 2 shows local neighborhoods of the four types of non-degenerate critical points. The criticality of p is characterized by the structure of the *oceans*, consisting of points x on the sphere around p with $f(x) < f(p)$, and *continents*, consisting of points x on the sphere with $f(x) > f(p)$. The *index* of a critical point is equal to the number of negative eigenvalues of the Hessian matrix. Therefore, minima, 1-saddles, 2-saddles, and maxima have indices equal to $0, 1, 2$, and 3, respectively.

Morse-Smale complex. An integral line of f is a maximal path in $\mathbb{M}$ whose tangent vectors agree with the gradient of f at every point of the path. Each integral line has a natural origin and destination at critical points of f where the gradient becomes zero. *Ascending* and *descending manifolds* are obtained as clusters of integral lines having common origin and destination respectively. The descending manifolds have dimension equal to the index of the destination and ascending manifolds have dimension equal to 3 minus the index of their origin. The descending manifolds of dimension 1 and 2 are called *descending arcs* and *descending disks*; the ascending manifolds of dimension 1 and 2 are called *ascending arcs* and *ascending disks*. The *Morse-Smale complex* partitions $\mathbb{M}$ into regions by clustering integral lines that share common origin and destination. For example, the 3D cells of the Morse-Smale complex cluster integral lines that originate at a given minimum and terminate at an associated maximum. The cells of different dimensions are called *crystals, quads, arcs,* and *nodes*. Note that the Morse-Smale complex is an overlay of ascending and descending manifolds, which individually partition $\mathbb{M}$ as well.
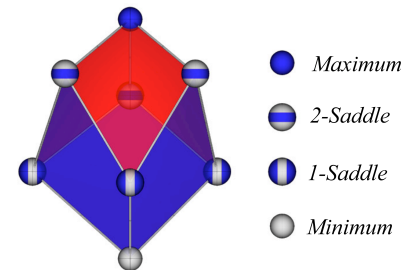

Figure 3: The boundary of a crystal in the Morse-Smale complex consists of lower-dimensional cells: quads, arcs, and nodes. Every crystal has a unique origin and destination node, the minimum and maximum, respectively, which are end points of integral lines lying within. Note that the glyphs for critical points are miniature versions of their local neighborhood, with shaded oceans and white continents. Quads contained in ascending disks are colored in red and those contained in descending disks are colored blue.
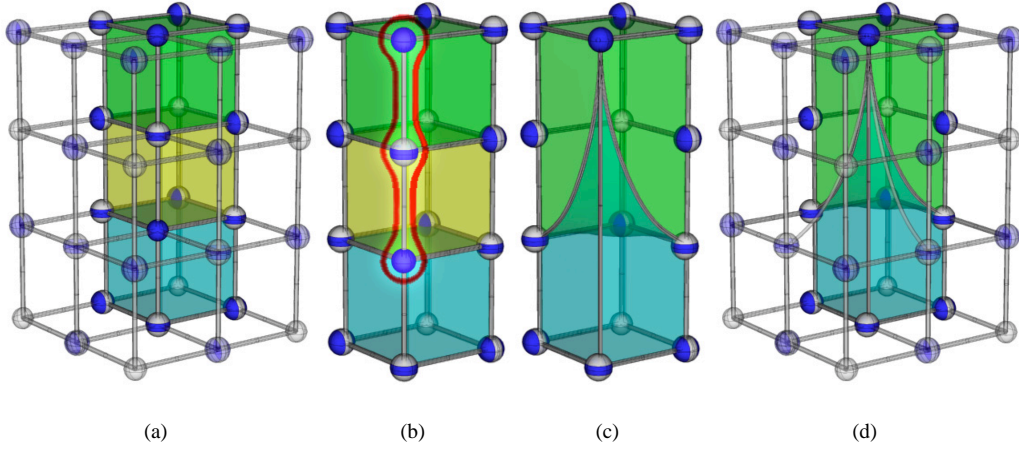
Figure 4: Snapshot of the Morse-Smale complex before and after a saddle-maximum cancellation. (a) Three of the twelve crystals affected by the cancellation are colored. (b) Closer view of the three crystals and the maximum-saddle-maximum triple that is merged into a single maximum. (c) After cancellation, all ascending arcs and disks originally flowing into the maximum below the saddle now flow into the maximum lying above the saddle. One crystal is deleted in this process. (d) Ascending arcs and disks in the other nine crystals are similarly re-routed resulting in a removal of three more crystals.

Cancellation. A minimal Morse function is generated from f by repeated cancellation of pairs of critical points. This operation is legal (*i.e.*, it can be realized by a local perturbation of the gradient vector field) if the indices of the two critical points differ by one and they are connected by a common arc in the Morse-Smale complex. For Morse functions defined on 3-manifolds, we three types of legal cancellations: minimum and 1-saddle, 1-saddle and 2-saddle, 2-saddle and maximum. Cancellations in Morse theory play a crucial role in proving important results, including the generalized Poincaré conjecture for higher dimensions [18]. Our use of cancellations is obvious: we use them to reduce the number of critical points and hence remove topological features. The local change in the Morse-Smale complex indicates a smoothening of the gradient vector field and hence a smoothening of the function f. The ordering of critical point pairs is given by the notion of persistence, which quantifies the importance of the topological feature associated with a pair. The persistence of a critical point pair is defined as the absolute difference in value of f between the two points.

Piecewise-linear functions. Scientific data is typically available as a discrete sample over a smooth manifold $\mathbb{M}$. The smooth manifold, in this case, is represented by a triangulation. Let K be a triangulation of the given 3-manifold $\mathbb{M}$. K consists of simplices of dimension 0,1,2, and 3, which we refer to as *vertices*, *edges*, *triangles*, and *tetrahedra*. The underlying space of K, $|K|$, consisting of all points in simplices of K, is homeomorphic to $\mathbb{M}$. The spherical neighborhood of a vertex v is represented by a triangulation of its vertex neighbors. Edges and triangles in this triangulation, called the *link* of v, are exactly the faces of tetrahedra that contain v. Let $f : |K| \rightarrow \mathbb{R}$ be a continuous function that is linear within each simplex of K. Assuming the value of f is given at the vertices, we linearly extend it within the simplex. Oceans consist of simplices in the link where f assumes values lower than $f(v)$. We count the number of components and tunnels in the ocean, using reduced Betti numbers, to distinguish regular from critical vertices and to classify the latter. Gradients, and hence integral lines, are not well defined for PL functions. However, monotonic curves and surfaces corresponding to arcs and quads of the Morse-Smale complex can be constructed by simulating a separation between integral lines that merge [7]. The function f has its critical points at the

nodes of this complex and is monotonic within all the arcs, quads, and crystals of the Morse-Smale complex. We work with this decomposition of K to obtain pairs of critical points that are to be canceled as the boundary nodes of a common arc.

3 SIMPLIFICATION

We simplify the Morse-Smale complex of a given Morse function f by performing a series of critical point pair cancellations. A cancellation simulates the smoothing of f by modifying gradient flows in the neighborhood of two critical points. Arcs connecting critical points are lines of steepest descent or ascent and therefore changing them affects the gradient flow behavior of the function. Rules that apply to gradient flow must be adhered to in the simplification process. For example, integral lines must remain disjoint.

Critical point pairs that we consider are end points of an arc in the Morse-Smale complex and therefore have consecutive indices. We group the pairs into two types: saddle-extremum (indices 1 and 0 or indices 2 and 3) and saddle-saddle (indices 1 and 2). The two types of cancellations are distinct in how they modify the gradient flow behavior of the function. The cancellation procedure is analogous to vertex removal in mesh simplification, with a pair of critical points being removed now instead of a single vertex, and the reconnection of the complex governed by rules of Morse theory rather than mesh geometry. For reasons of clarity, we illustrate the two types of cancellations using prototypical figures of Morse-Smale complexes. The description, however, holds for all possible configurations.

Saddle-extremum cancellation. The saddle-extremum cancellation removes either a 2-saddle-maximum pair or a 1-saddle-minimum pair. Clearly, the two pairs are dual to each other as can be seen by negating the function: maxima become minima, 2-saddles become 1-saddles and vice-versa. We restrict our discussion below to 2-saddles and maxima. A 2-saddle, by definition, is connected by ascending arcs to exactly two maxima. When one of these maxima is removed in a saddle-maximum cancellation, integral lines ending at this maximum flow toward the second maximum. We think of a saddle-maximum cancellation as a merging of three critical points into one maximum. Applying the

saddle-maximum cancellation simplifies the function by removing a "bump". Figure 4 shows how the integral lines terminating at the two maxima flow into the existing maximum after cancellation.

The saddle-maximum cancellation is similar to its two-dimensional analog, which is also implemented as a merging of three critical points. We merge neighboring cells in the ring around the saddle-maximum arc to reconnect the complex. Therefore, besides removing two critical points, this cancellation also removes several crystals, quads, and arcs from the complex.

A saddle-maximum cancellation is legal only if the 2-saddle is connected to two distinct maxima. If this condition is not met, then we recognize that the cancellation causes a *strangulation* of the descending disk that originates at the 2-saddle. Indeed, it is not possible to route the integral lines terminating at the 2-saddle if we do cancel such a saddle-maximum pair. Figure 5 shows this configuration.
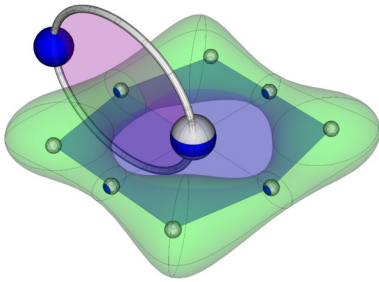

Figure 5: The two integral lines beginning at a 2-saddle flow to the same maximum. Canceling the saddle-maximum pair causes a strangulation of the blue descending disk because integral lines terminating at the 2-saddle are left without a destination.

Saddle-saddle cancellation. The saddle-saddle cancellation removes a 1-saddle-2-saddle pair. This cancellation does not have an analog in lower dimensions and therefore intuition from 2D does not help. A 1-saddle descends to exactly two minima, and a 2-saddle ascends to exactly two maxima. After canceling this saddle pair, we need to ensure that the two pairs of extrema originally separated by these saddles remain that way. This necessitates the introduction of new cells to fill in space between the two pairs of extrema. The easiest way to think about this cancellation is to consider what happens to the descending disk originating from the 2-saddle, and the ascending disk originating from the 1-saddle. Upon cancellation, these two disks disappear and neighboring disks stretch out and share their boundary. Figure 6 illustrates the operation by showing the descending disks before and after cancellation. We can no longer consider the cancellation as a merging of three critical points, as we did for the saddle-maximum cancellation. Consider the descending disk that is removed by the cancellation. The boundary of this disk consists of alternating 1-saddles and minima. Arcs lying within the disk connect the source 2-saddle to 1-saddles on the boundary. One of these 1-saddles is involved in the cancellation. This 1-saddle and its two descending arcs are deleted by the cancellation. Descending disks that contain the two deleted descending arcs in their boundary expand to share the boundary of the removed disk. Similarly, one ascending disk is removed and its boundary is shared by the neighboring ascending disks.

The best way to think about how to reconnect the complex after a saddle-saddle cancellation is, again, in terms of ascending and descending disks. All surviving descending disks expand to share the
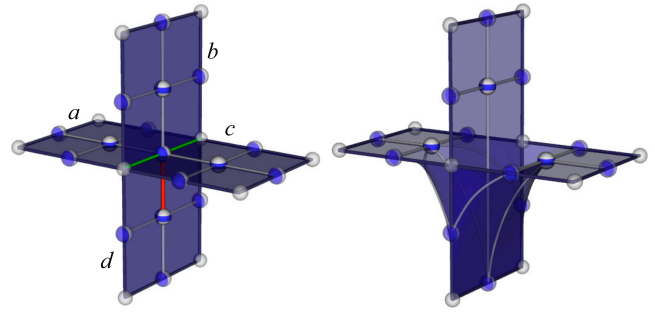

Figure 6: A saddle-saddle cancellation: the figure on left shows the relevant descending disks before the cancellation. The red arc connects the pair to be canceled. All four disks (a,b,c, and d) have two common descending arcs (shown in green) on their boundary, both originating from the 1-saddle to be removed. The figure on right shows the descending disks that remain after cancellation. The green descending arcs are deleted from the boundary of the three surviving disks, which now extend and inherit the boundary of d.

boundary of the deleted disk thereby creating connections between surviving 2-saddles and 1-saddles on the newly inserted boundary. Similarly, surviving 1-saddles connect to 2-saddles on the newly inserted boundary of their ascending disks. This gives the full reconnectivity of the complex after a cancellation.

The Morse-Smale complex actually gains cells after a saddle-saddle cancellation because re-routing the descending disks creates new intersections between ascending and descending disks. Figure 7 shows the cells destroyed and created by this operation. For simplicity, these figures show only three of the twelve crystals destroyed by the cancellation and the cells that reconnect this complex within this region. Figure 8 shows how one of these newly created crystals fills in the space occupied by the destroyed crystals. Introducing new cells is counter-intuitive because simplification actually does not lead to a smaller complex. Although the Morse-Smale complex apparently gains complexity in size, the function is smoothed by the removal of these saddle pairs. Also, note that the cells created by the saddle-saddle cancellation are introduced into rings around saddle-extremum pairs. A future saddle-extremum cancellation will remove all these cells leading to a smoother Morse function and a smaller Morse-Smale complex. A simple proof that the algorithm terminates, despite the increase in number of cells following the saddle-saddle cancellation, follows from the fact that every cancellation results in the removal of a pair of nodes from the complex. In theory, the maximum number of cancellations is bounded by half the number of critical points. In practice, however, we will stop earlier to preserve the more persistent features. The saddle-saddle cancellation introduces significant complexity that is not found in lower dimensions. Furthermore, the complexity of this cancellation indicates that future extensions into higher dimensions will be non-trivial.

An arc connecting two saddles does not guarantee a valid saddle-saddle cancellation. Performing a cancellation could lead to the formation of a *pouch*. This situation occurs when a crystal incident on the arc has exactly two quads, one connecting the saddle pair to a minimum and the other connecting them to a maximum. Removing the 1-saddle and 2-saddle creates a crystal with zero saddles and zero quads corresponding to a possibly valid Morse function but resulting in an invalid combinatorial structure for the Morse-Smale complex.

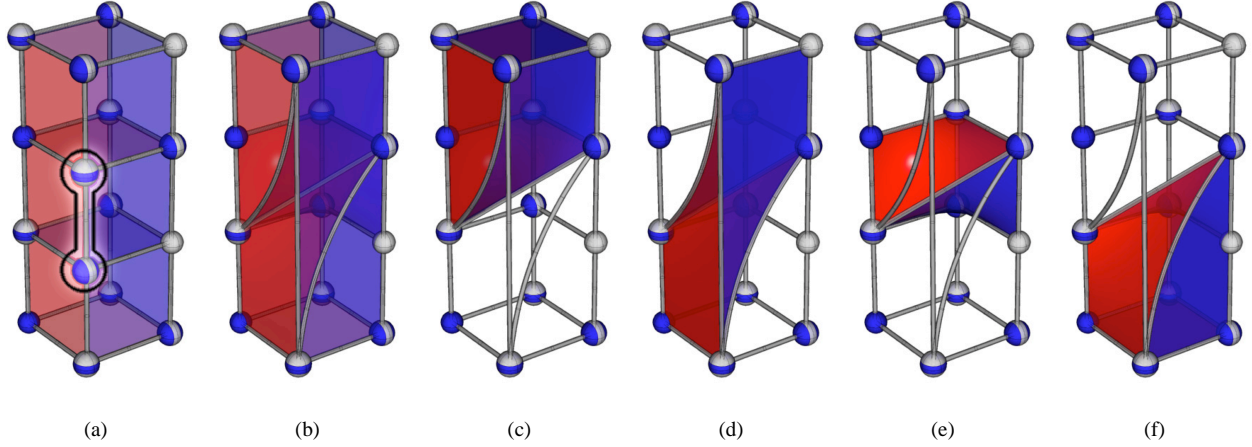(a)   (b)   (c)   (d)   (e)   (f)

Figure 7: Cells destroyed and created by a saddle-saddle cancellation. The neighborhood of the saddle-saddle pair is divided into four regions of 3 crystals each and only one region is shown for reasons of clarity. Cells within the other regions are modified in a similar manner. (a) Three crystals before cancellation of the highlighted pair of saddles. (b) Four crystals after cancellation . (c) The top crystal stretches down to the lower 2-saddle. (d) a crystal defined by a new minimum-maximum pair stretches from the lower left to the upper right corner of the region. (e) The middle crystal shrinks after losing the saddle-saddle pair. (f) The bottom crystal stretches up to the 1-saddle on the upper right.
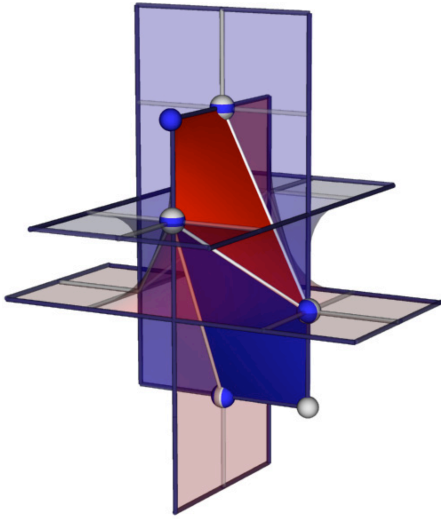

Figure 8: A new crystal introduced by the saddle-saddle cancellation. The two quads in brighter red and blue colors lie on the face of the newly introduced crystal.

4   IMPLEMENTATION

Storage. We follow a minimalist approach to store the Morse-Smale complex during the simplification process. Critical points and arcs are stored as a multi-graph. An arc connects two critical points whose indices differ exactly by 1. The critical points are stored in an array. We use a skip list to store the edges, because we require efficient access both in sorted as well as random order while allowing for fast insertions and deletions. We do not store quads and crystals of the complex explicitly because the one-dimensional representation is all we need to determine legal cancellations and to perform the cancellation. Moreover, quads and crystals can be recovered at runtime from our multi-graph whenever needed.

We maintain a list of arcs ordered by the persistence value of the corresponding pair of critical points. The arc having lowest persistence is removed from this list and its end points are chosen for cancellation. If this cancellation is recognized as being illegal, then we discard the arc and choose the next arc from our sorted list. A simplification is represented using the set of arcs that are removed from the complex, the set of new arcs that are created in the complex, and the pair of critical points that are canceled. Defining the simplification in such a way is important to enable forward and backward navigation of the simplification sequence. An anti-cancellation is implemented by swapping the sets representing arcs that were removed and arcs that were created, and the critical point pair is re-inserted into the complex. Insertion and deletion of arcs and critical points is implemented as a simple activation and deactivation of simplices.

Pre-processing. Given a PL function defined on a volumetric mesh, we first construct an artificial complex by inserting dummy critical points at the barycenters of all simplices with the index of criticality being equal to the dimension of the simplex. All input data points now become local minima. We add new arcs connecting the barycenter of a simplex with the barycenters of its faces. The dummy vertices and arcs naturally subdivide each simplex into arcs, quads, and crystals. Minima inherit the function value from the corresponding data point. The dummy nodes are assigned function values so that all dummy arcs have zero persistence. Initial cancellations performed on this artificial complex removes the dummy nodes and arcs resulting in a complex that represents features of the input data. A barycentric subdivision when applied to a cube also gives an artificial complex with minima at the input data locations. We begin with this artificial complex when the input data is available over a cube lattice. Figure 9 illustrates how we subdivide a tetrahedron and a cube. We choose to construct the complex this way because of its simplicity as opposed to the intricate algorithm described by Edelsbrunner et al. [7]. A disadvantage of starting with this artificial complex is that the number of nodes is equal to the total number of simplices in the input, thereby limiting the size of the dataset that can be efficiently processed. However, all dummy nodes are removed in a pre-processing step and therefore exploration of the data is done interactively.

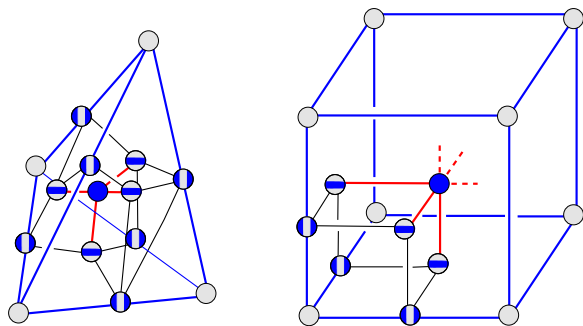Besides its simplicity, another advantage of starting with the ar-

Figure 9: Creation of an artificial complex by subdividing tetrahedra or cubes: dummy critical points are introduced thereby converting original data points into local minima. Function value at dummy critical points is infinitesimally larger than input data points and therefore have very low persistence values.

tificial complex is that it automatically resolves the issue of degenerate critical points. *Multiple saddles*, which are typically unfolded into simple 1- and 2-saddles, are not present in the artificial complex. This is because each 1-saddle, at the barycenter of an edge, now has exactly two descending arcs connecting it to minima at the end points of the edge. Similarly, all 2-saddles are simple because there are exactly two ascending arcs that connect each one of them to local maxima. Both types of cancellation do not change the number of extrema that connect to a saddle. Therefore, no new multiple saddles are introduced. Multiple saddles do exist in typical input data but our construction forces the splitting of these into simple saddles.

Boundary. Our input data is defined over a volumetric domain in $\mathbb{R}^3$ that has a non-empty boundary. We use a standard technique from point-set topology, called *One-point compactification*, to convert the domain into a 3-manifold. The compactification involves addition of a vertex at infinity that connects to all boundary vertices. This extension of the triangulation K is a simple and efficient way to handle the domain boundary and has been described before by several authors [8, 16, 20]. Instead of explicitly adding the vertex at infinity and simplices connecting it to the boundary, we create a layer around the domain consisting of dummy critical points for each simplex that contains the vertex at infinity. These dummy critical points become nodes of the artificial complex and are removed when we cancel zero-persistence pairs. We restrict all cancellations to pairs that lie completely in the interior of the domain or within the boundary to ensure that we do not change the topology of the domain.

5 RESULTS

We pre-process the input data by first creating an artificial complex and then removing all dummy critical points by canceling zero-persistence pairs. We perform further simplification of the complex in an interactive process to identify features. The artificial complex that we create has a large number of critical points and is not a succinct representation of the data. After the dummy critical points are removed, the complex provides an efficient representation of features in the data. We identify important features as regions associated with persistent critical points. Similar to Takahashi et al [20], we automatically design a transfer function to enhances critical values that correspond to these features. Our simplification allows us to limit the number of critical values affecting the transfer function to exactly those representing important features.

Feature identification. We show that our simplification technique extracts known features and removes noise in well-studied datasets. Table 1 lists the data sets used, their sizes, the number of initial critical points present in the dataset, and time taken for pre-processing. After pre-processing, all critical pairs with non-zero persistence can be removed at the rate of 1200 cancellations per second for all data sets. All experiments were performed on a desktop PC with 2GB RAM and an Intel Xeon processor running at 3GHz. The first data set contains electron density distribution in a C_4H_4 molecule. After zero-persistence critical point pairs are removed, the complex correctly outlines the bond structure of the C_4H_4 molecule. High-persistent maxima correspond exactly to locations of atoms in the molecule and ascending arcs connecting 2-saddles and these maxima correspond to bonds between atoms. (see Figure 10). This correspondence is a visual depiction of the topological approach to identifying atoms in molecules as proposed by the AIM theory [1].

Data set	Size	n_{crit}	p_{max}	t_{pre}
C_4H_4	$33 \times 33 \times 33$	95	23	2m 21s
Fuel	$64 \times 64 \times 64$	219	255	18m 18s
Hydrogen	$64 \times 64 \times 64$	669	78	21m 42s
Neghip	$64 \times 64 \times 64$	1045	255	16m 38s
Bonsai	$32 \times 32 \times 32$	5053	255	2m 15s

Table 1: Datasets used in experiments, their size, number of critical points (n_{crit}), persistence value of largest feature (p_{max}), and timing results for pre-processing (t_{pre}).
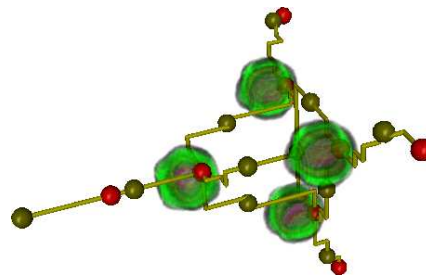

Figure 10: Atoms and bonds in the C_4H_4 molecule are identified by high-persistence critical points and ascending arcs in the simplified Morse-Smale complex.

The remaining data sets in Table 1 are obtained either from simulations or from MRI / CT scans. Visualization of our results for these data sets are shown in Figures 1, 11, 12 and 13. The hydrogen atom dataset gives the spatial electron density in a hydrogen atom subjected to a large magnetic field. The dataset consists of high density around the nucleus, two regions of high density on either side, and a torus of high density around the nucleus. We correctly identify these features. After the zero-persistence critical points are removed, there are still two disks of saddles separating the maxima in the dataset, which corresponds to noise. Initial cancellations remove the spurious 1-saddles and 2-saddles, leaving behind four maxima, which represent the four regions of high electron density. Distinctive features in the fuel, neghip, and bonsai data sets are revealed using a low threshold. In our experiments, we observe that a threshold value equal to 10% of the maximum persistence is able to detect and remove all insignificant features. This is most clearly seen in the neghip data set, where we are able to isolate the different clusters of atoms automatically. For the bonsai data set, we also see that the shape of the tree is traced by ascending arcs of the simplified Morse-Smale complex.

Noise removal. We use synthetic data to illustrate how the cancellation of critical point pairs is useful for noise removal and hence a robust identification of features. We superimpose cosine functions with different magnitudes and frequencies to generate a noisy input:

$$
\begin{aligned}
f(x,y,z) &= \cos(\alpha x) + \cos(\alpha y) + \cos(\alpha z) \\
&+ 0.5\,(\cos(4\alpha x) + \cos(4\alpha y) + \cos(4\alpha z)) \\
&+ 0.25\,(\cos(8\alpha x) + \cos(8\alpha y) + \cos(8\alpha z)).
\end{aligned}
$$

The high frequency components simulates noise whereas low frequency component are features that we wish to identify. Figure 14 shows how the function is successively smoothed by removing critical points based on persistence. High frequency noise with lower amplitude is removed first followed by the next tier of noise, leaving behind the sinusoidal function.

6 CONCLUSIONS

We have described an algorithm to simplify a three-dimensional Morse function by canceling pairs of critical points in its Morse-Smale complex and demonstrated its application to the identification of features in volumetric scalar fields. This topological approach supports a direct manipulation of features including their detection, ordering, and removal during an interactive exploratory study. The combinatorial nature of our algorithm leads to robust and efficient implementations and hence allows us to perform topological analysis on complex and noisy data sets. We use the notion of persistence to determine the importance of critical points and the features that they represent. This measure has been proven to be robust under the presence of noise [5]. Other spatial measures have also been successfully used to detect important topological features [3]. We plan to incorporate such measures into our framework and perform a comparative study. Computing the Morse-Smale complex remains a bottleneck in our computations. We are exploring alternate methods to compute the Morse-Smale complex for large data sets. Future work also includes designing a multiresolution data structure for three-dimensional continuous functions by performing independent cancellations in the Morse-Smale complex. A numerical realization of the cancellations discussed in this paper is a necessary and nontrivial ingredient in the design of a multiresolution data structure. In particular, an efficient realization of the saddle-saddle cancellation poses a challenging problem.

REFERENCES

[1] BADER, R. F. W. *Atoms in Molecules.* Clarendon Press, Oxford, 1995.

[2] BREMER, P.-T., EDELSBRUNNER, H., HAMANN, B., AND PASCUCCI, V. A topological hierarchy for functions on triangulated surfaces. *IEEE Transactions on Visualization and Computer Graphics 10*, 4 (2004), 385–396.

[3] CARR, H., SNOEYINK, J., AND VAN DE PANNE, M. Simplifying flexible isosurfaces using local geometric measures. In *Proc. IEEE Visualization* (2004), pp. 497–504.

[4] CIGNONI, P., CONSTANZA, D., MONTANI, C., ROCCHINI, C., AND SCOPIGNO, R. Simplification of tetrahedral meshes with accurate error evaluation. In *Proc. IEEE Conf. Visualization* (2000), pp. 85–92.

[5] COHEN-STEINER, D., EDELSBRUNNER, H., AND HARER, J. Stability of persistence diagrams. In *Proc. 21st Sympos. Comput. Geom.* (2005), pp. 263–271.

[6] DE LEEUW, W., AND VAN LIERE, R. Collapsing flow topology using area metrics. In *Proc. IEEE Conf. Visualization* (1999), pp. 149–154.

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

[8] EDELSBRUNNER, H., HARER, J., AND ZOMORODIAN., A. Hierarchical Morse-Smale complexes for piecewise linear 2-manifolds. *Discrete and Computational Geometry 30*, 1 (2003), 87–107.

[9] EDELSBRUNNER, H., LETSCHER, D., AND ZOMORODIAN., A. Topological persistence and simplification. *Discrete and Computational Geometry 28*, 4 (2002), 511–533.

[10] GARLAND, M., AND HECKBERT, P. S. Surface simplification using quadric error metrics. In *Proc. SIGGRAPH* (1997), pp. 209–216.

[11] GARLAND, M., AND ZHOU, Y. Quadric-based simplification in any dimension. *ACM Transactions on Graphics 24*, 2 (2005), to appear.

[12] GUSKOV, I., AND WOOD, Z. Topological noise removal. In *Proc. Graphics Interface* (2001), pp. 19–26.

[13] HOPPE, H. Progressive meshes. In *Proc. SIGGRAPH* (1996), pp. 99–108.

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

[15] MILNOR., J. *Morse Theory.* Princeton Univ. Press, New Jersey, 1963.

[16] NATARAJAN, V., AND EDELSBRUNNER, H. Simplification of three-dimensional density maps. *IEEE Transactions on Visualization and Computer Graphics 10*, 5 (2004), 587–597.

[17] REEB, G. 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] SMALE, S. Generalized Poincaré's conjecture in dimensions greater than four. *Ann. of Math. 74* (1961), 391–406.

[19] SMALE, S. On gradient dynamical systems. *Ann. of Math. 74* (1961), 199–206.

[20] TAKAHASHI, S., TAKESHIMA, Y., AND FUJISHIRO, I. Topological volume skeletonization and its application to transfer function design. *Graphical Models 66*, 1 (2004), 24–49.

[21] TRICOCHE, X., SCHEUERMANN, G., AND HAGEN, H. Continuous topology simplification of planar vector fields. In *Proc. IEEE Conf. Visualization* (2001), pp. 159–166.

[22] WOOD, Z., HOPPE, H., DESBRUN, M., AND SCHRÖDER, P. Removing excess topology from isosurfaces. *ACM Transactions on Graphics 23*, 2 (2004), 190–208.
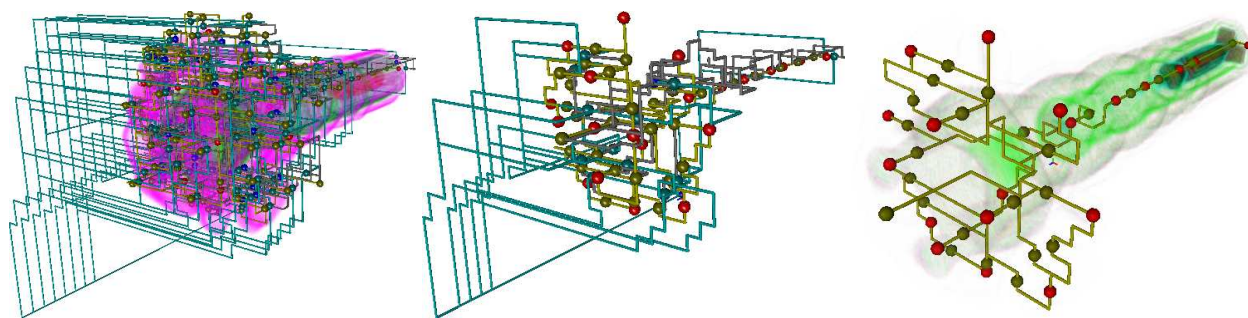
Figure 11: Features extracted from the fuel dataset (maximum persistence = 255): the input data is shown on the left. The middle figure shows the Morse-Smale complex after performing cancellations that remove all critical points with persistence value lower than 1. The figure on right shows features represented by critical points with persistence value greater than 7.4.
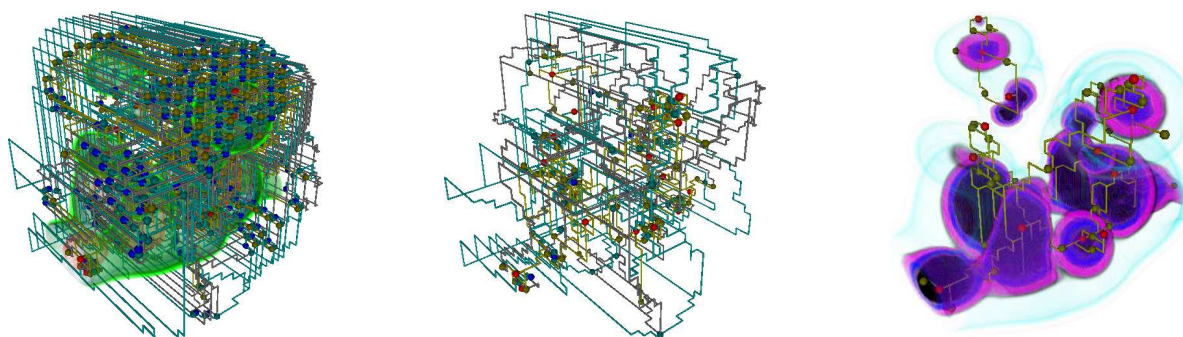

Figure 12: Features extracted from the neghip dataset (maximum persistence = 255): the input data has over a thousand critical points, many of which have low persistence values and are removed. The middle figure shows the complex after all critical points with persistence value lower than 1 are canceled. Canceling critical points with persistence value lower than 36 isolates the various clusters of atoms present in this protein.
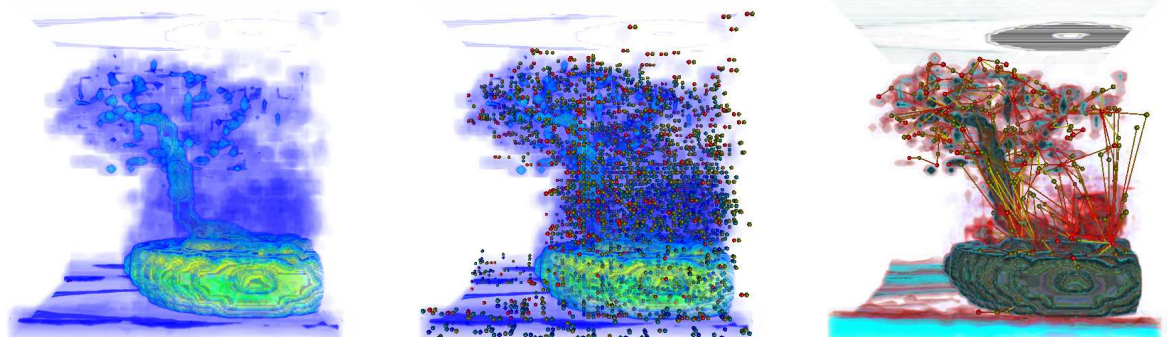

Figure 13: Features extracted from the bonsai tree data set (maximum persistence = 255): topology-based simplification applied to the CT scan of a bonsai tree identifies important features. The input data (left) is a down sampled version and has noisy regions which manifest as clusters of critical points (middle). Removing all critical points with persistence value lower than 19 makes these regions smooth resulting in a better identification of features using volume rendering (right).
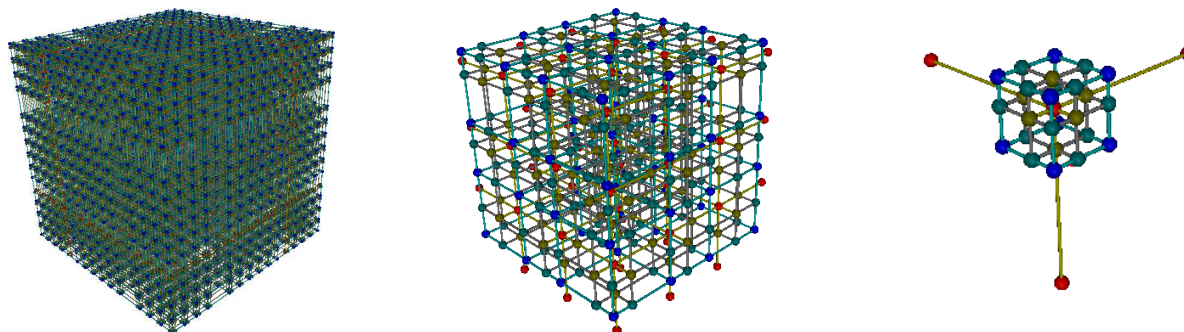

Figure 14: High frequency noise is detected as a set of features with low persistence and removed by canceling pairs of critical points. Left: the Morse-Smale complex of a noisy sinusoidal function. Middle and right: high frequency noise is removed after successive stages of simplification.