PURPOSE-LED
PUBLISHING™

Topological feature extraction and tracking

View the article online for updates and enhancements.

Topological feature extraction and tracking

P-T Bremer, E M Bringa, M A Duchaineau, A G Gyulassy, D Laney, A Mascarenhas and V Pascucci

Lawrence Livermore National Laboratory, Livermore, CA

E-mail: {bremer5, mascarenhas1}@llnl.gov

Abstract. Scientific datasets obtained by measurement or produced by computational simulations must be analyzed to understand the phenomenon under study. The analysis typically requires a mathematically sound definition of the features of interest and robust algorithms to identify these features, compute statistics about them, and often track them over time. Because scientific datasets often capture phenomena with multi-scale behaviour, and almost always contain noise the definitions and algorithms must be designed with sufficient flexibility and care to allow multi-scale analysis and noise-removal. In this paper, we present some recent work on topological feature extraction and tracking with applications in molecular analysis, combustion simulation, and structural analysis of porous materials.

1. Introduction

With rapidly increasing computational power scientists are now able to perform simulations of many fundamental physical processes at increasing resolution and complexity [8, 7]. The resulting data must be analyzed to understand the underlying phenomenon. There is a well-developed research area of visualization and data analysis that addresses this issue and any survey of the results is beyond the scope of this paper; the interested reader can begin at [17, 11].

This paper describes recent developments in using concepts from topology [15], and Morse theory [13] to develop robust combinatorial algorithms for feature extraction and tracking. We emphasize the power of these methods by describing their application in several data analysis tasks ranging from protein structure analysis to the study of channel structures in porous materials.

2. Mathematical Background

We need some background from Morse theory [12, 13] and from combinatorial and algebraic topology [1, 15].

Smooth maps on manifolds. Let $f \colon \mathbb{M} \to \mathbb{R}$ be a smooth map. A point $x \in \mathbb{M}$ is a *critical point* of f if the gradient of f vanishes at x, and the value $f(x)$ is a *critical value.* Non-critical points and non-critical values are called *regular points* and *regular values*, respectively. A critical point x is *non-degenerate* if the Hessian (matrix of second-order partial derivatives) at x is non-singular. The *index* of a critical point x, denoted by index x, is the number of negative eigenvalues of the Hessian. For $d = 3$ there are four types of non-degenerate critical points: the *minima* (index 0), the *1-saddles* (index 1), the *2-saddles* (index 2), and the *maxima* (index 3). A function f is *Morse* if all critical points are non-degenerate with distinct values.

Morse-Smale complex. An *integral line* is a maximal path on $\mathbb{M}$ whose tangent vectors agree with the gradient of f. The *ascending manifold* of a critical point x is the union of x and all integral lines that *start* at x. The *descending manifold* of x is defined symmetrically as the union of a critical point x and all integral lines that *end* at x. One can superimpose the ascending and descending manifolds of all critical points to create the *Morse-Smale complex* (or MS complex) of f [6, 2], see Figure 1(a)–(d). The *nodes* of this complex are the critical points of f, its *arcs* are integral lines starting or ending at saddles and its *regions* are the non-empty intersections of ascending and descending 2-manifolds. More details on the definition of the MS-complex on 2-manifold triangle meshes and algorithms to compute it can be found in Bremer et al. [2].
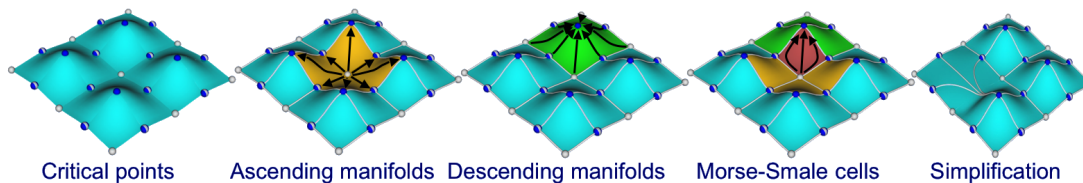

Figure 1: MS-complex construction, simplification and topologically valid approximation: (a) Morse function with critical points shown; (b) Ascending manifolds; (c) Descending manifolds; (d) MS-complex; (e) MS-complex and manifold after simplification. Maxima are solid blue, minima are solid white, and saddles are mixed.

Simplification. It is often useful to simplify an MS-complex to remove noise and to perform multi-scale function analysis. Following [2], we perform *cancellations* of arc-connected critical point pairs to simplify an MS-complex. Cancellations are ranked by their *persistence* —the absolute difference in function value between the cancelled critical point pair. Figures 1(e) and (f) shows an example of a topological simplification and a corresponding approximation of f.

Computation. In practice, one usually deals with piece-wise linear (PL)-functions given at the vertices of a triangulation. See [2, 3] for a detailed discussion on how to translate concepts from the generic smooth functions discussed above to PL-functions.

Jacobi Sets. We define Jacobi set of two Morse functions, $f, g \colon \mathbb{M} \to \mathbb{R}$; the general definition is found in [5]. For a regular value $t \in \mathbb{R}$, we have the level set $g^{-1}(t)$ and the restriction of f to this level set, $f_t : g^{-1}(t) \to \mathbb{R}$. The *Jacobi set* of f and g is the closure of the set of critical points of the functions f_t, for all $t \in \mathbb{R}$. The closure operation adds the critical points of f restricted to level sets at critical values, as well as the critical points of g, which form singularities in these level sets. Figure 2 illustrates the definition by showing the Jacobi set of two smooth functions on a portion of the two-dimensional plane. The

Figure 2: Solid and dotted level curves represent functions f and g, respectively. The Jacobi set is the bold solid curve.

Jacobi set of f and g may consist of several components, and in the assumed generic case each is a closed 1-manifold. In Figure 2, function g is a ramp that grows from left to right and f_t can be considered a time-varying function defined on a line segment. The Jacobi set tracks the critical points of f_t for continuously changing values of t.
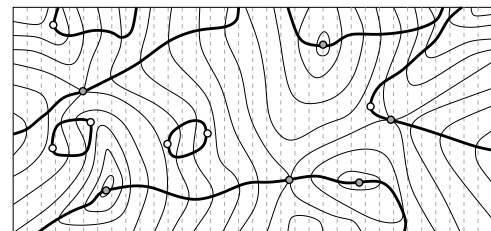
3. Feature Extraction: Segmentation, Simplification, and Hierarchies

We describe how the topological concepts and algorithms described in Section 2 are applied in practical applications ranging from the analysis of proteins to the study of core structures in high velocity impact simulations.
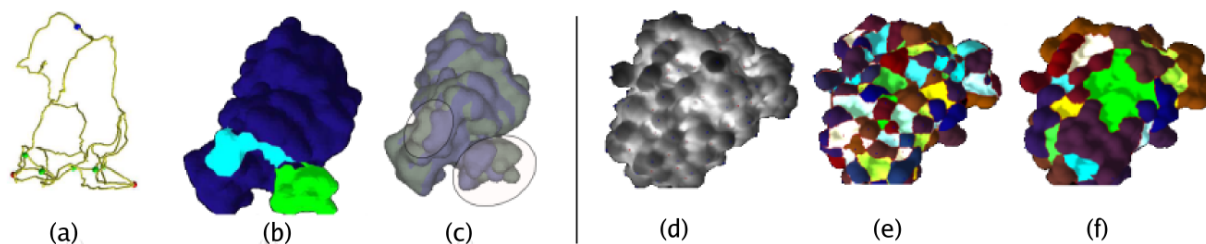
Figure 3: Left, detecting semi-rigid components in a protein. (a) Simplified MS-complex with one minimum (blue), two maxima (red), and four saddles (green). (b) Segmentation based on the simplified MS-complex. (c) Corresponding semi-rigid components. Right, detecting cavities and protrusions. (d) The atomic density function: Darker regions correspond to protrusions and lighter regions correspond to cavities. (e) and (f) Simplified triangulations and their segmentations with approximately 300 and 80 segments, respectively.

3.1. Analysis of protein-protein interaction

In Natarajan et al. [16], we describe a method for segmentation of molecular surfaces to study semi-rigid structures and the role of cavities and protrusions in protein-protein interactions. We use the *skin surface* [4] representation of the protein define a relevant function on this surface and compute the Morse-smale decomposition of the surface. They use this decomposition in two applications:

a. Detecting semi-rigid structures. Semi-rigid structures are important in the study of the so called *hinge-motion* which occurs when proteins interact. The function for segmentation is obtained by aligning two conformations of the molecule and computing the distance between each point in one conformation to its nearest neighbor in the other conformation. Figure 3 left, shows the simplified MS-complex of this function, and the corresponding semi-rigid components which are regions associated with a maximum.

b. Detecting cavities and protrusions. Protein interactions typically occur at binding sites which are geometrically complimentary. The study of cavities and protrusions on the surfaces of proteins is therefore important to detect potential binding sites. Natarajan et al. define a variant of the *atomic density function* originally defined in [14]. As shown in Figure 3 right, the MS-complex is able to detect protrusions and cavities, and the simplification hierarchy allows the user to study the molecule at different scales.

3.2. Core structures in porous materials

The study of structural properties of materials under stress is important in design and manufacturing applications. In Gyulassy et al. [9], we use the MS-complex to analyze stable channel structures from an atomistic simulation of a porous solid under impact with a high density projectile. We present two methods to construct the Morse-Smale complex. The first method constructs a standard distance field from a given interface surface embedded within the domain, but may produce many spurious critical points slowing down the MS-complex computation. The second method creates a better distance field by advancing a front from the original surface in a controlled manner, ensuring that the field created is topologically clean leading to a fast
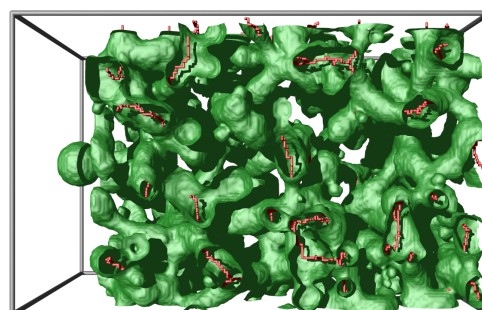

Figure 4: Channel structures in porous media. The green iso-surface separates solid material and empty space. The red curves connecting maxima and 2-saddles represent channel structure cores.

MS-complex computation. Figure 4 shows the interface between solid material and empty space and the core of the channel structures.

4. Feature tracking

Temporal tracking of features in simulation data is important to understand how the phenomenon under study evolves over time. In this section, we briefly describe two techniques for tracking topological features. The first employs Jacobi sets to track critical points over time, the second uses distance measures on graphs to track channel structures in a porous medium.

Tracking critical points with Jacobi sets. We can defined Jacobi sets for a time-varying function $f : \mathbb{M} \times \mathbb{R} \to \mathbb{R}$, by introducing an auxiliary function for time as $g : \mathbb{M} \times \mathbb{R} \to \mathbb{R}$ defined by $g(x,t) = t$. From the general definition of the Jacobi set, the level set $g^{-1}(t) = \mathbb{M} \times t$, and the restriction of f to this level set is $f_t : \mathbb{M} \times t \to \mathbb{R}$. The Jacobi set then defines the trajectory of the critical points of f_t with varying time.

We implement the algorithm described in Edelsbrunner and Harer [5] to compute the Jacobi set for piecewise-linear functions. We first connect the sample points into a simplicial complex and linearly interpolate the function values within each simplex to define the function for space-time. Figure 5(a) shows the Jacobi set for a 2D combustion simulation dataset. The full-resolution
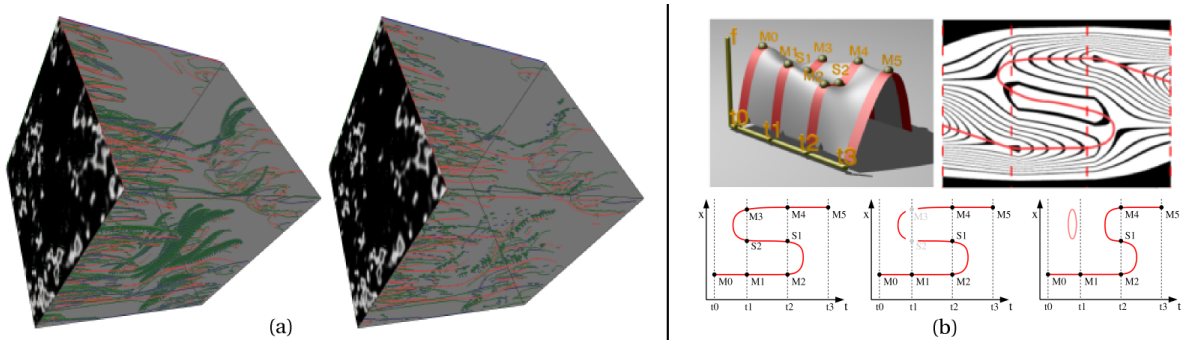

Figure 5: (a) Jacobi set of 2D combustion simulation. Full-resolution (left), simplified (right). The textured quad shows the function at time $t = 0$, with time increasing left to right within each cube. Maxima are red, saddles are green, and minima are blue. (b) Jacobi set simplification. Top left, a time-varying 1D function. Top right, view of level-sets and the Jacobi set (red). Bottom left, the Jacobi set with maxima and minima marked at discrete time-steps. Bottom middle, cancelation of a critical point pair. Bottom right, removal of a small loop.

Jacobi set (Figure 5(a) left) contains dense noisy regions which we identify and clean to obtain the simplification (Figure 5(a) right). As shown in Figure 5(b), we use the Morse-Smale complex of f_t at discrete time-steps to pair critical points, cancel pairs below a persistence threshold, and remove small loops of the Jacobi set that lie entirely within successive time-steps.

Tracking channel structures in a porous solid. In [9], we track the channel structures in a porous medium under impact for successive time-steps of the simulation to study how the impact affects the material. The tracking is performed by computing a distance measure between each point on the graph structure of the channel at time t with the graph at time $t + 1$. We use several different criteria to evaluate the tracking including the two-sided Hausdorff distance, and the average minimum distance. Figure 6 shows a visualization of three tracking operations at successive pairs of time-steps 500, 12750, 25500, and 51000. The red and blue lines show the displacement arcs between corresponding arcs. The displacements correlate well with the fact
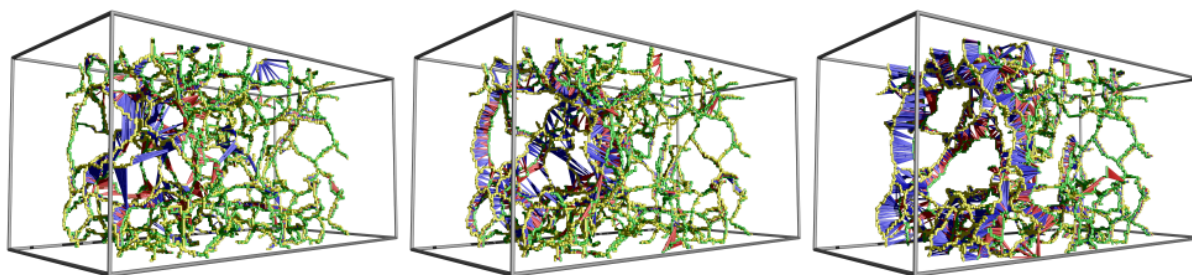
Figure 6: Tracking core structures between time-steps 500 and 12750 (left), 12750 and 25500 (middle), 5500 and 51000 (right). Red and blue lines indicate regions of greater displacement.

that the projectile impacts the solid on the left face and the crater expands towards the right, and that it is localized because the medium absorbs the shock wave efficiently.

5. Conclusion

We present recent research on topological feature extraction and tracking applied to the analysis of scientific datasets. Using notions from smooth Morse theory adapted to piecewise-linear functions we have developed a suit of techniques that allow us to: mathematically define features in a large variety of different applications; robustly extract these feature even from very large data sets; and organize them into a multi-scale hierarchy with guaranteed error bounds. In the future, we aim to extend our techniques to higher dimensional scalar functions and vector fields.

[1] P. S. Alexandrov, *Combinatorial topology*, Dover, Mineola, NY, USA, 1998.
[2] P.-T. Bremer, H. Edelsbrunner, B. Hamann, and V. Pascucci, *A topological hierarchy for functions on triangulated surfaces*, IEEE Trans. on Visualization and Computer Graphics **10** (2004), no. 4, 385–396.
[3] P.-T. Bremer and V. Pascucci, *A practical approach to two-dimensional scalar topology*, 2006, p. to appear.
[4] H.-L. Cheng, T. K. Dey, H. Edelsbrunner, and J. Sullivan, *Dynamic skin triangulation*, Symposium on Discrete Algorithms, 2001, pp. 47–56.
[5] H. Edelsbrunner and J. Harer, *Jacobi sets of multiple morse functions*, Foundations of Computational Mathematics (2004), 37–57.
[6] H. Edelsbrunner, J. Harer, and A. Zomorodian, *Hierarchical Morse-Smale complexes for piecewise linear 2-manifolds*, Discrete Comput. Geom. **30** (2003), 87–107.
[7] A. A. Mirin et al., *Very high resolution simulation of compressible turbulence on the ibm-sp system*, Supercomputing '99 **00** (1999), 70.
[8] F. Gygi et al., *Large-scale electronic structure calculations of high-z metals on the bluegene/l platform*, Supercomputing '06 (New York, NY, USA), ACM Press, 2006, p. 45.
[9] A. Gyulassy, M. Duchaineau, V. Natarajan, V. Pascucci, E. Bringa, A. Higginbotham, and B. Hamann, *Topologically clean distance fields*, IEEE Transactions on Visualization and Computer Graphics (to appear) (2007), xxx—xxx.
[10] C. Johnson, *Top scientific visualization research problems*, IEEE Computer Graphics and Applications **24** (2004), no. 4, 13–17.
[11] C. Johnson and C. Hansen, *Visualization handbook*, Academic Press, Inc., Orlando, FL, USA, 2004.
[12] Y. Matsumoto, *An introduction to morse theory. translated from japanese by k. hudson and m. saito*, American Mathematical Society, 2002.
[13] J. W. Milnor, *Morse theory*, Princeton Univ. Press, New Jersey, USA, 1963.
[14] J. C. Mitchell, R. Kerr, and L. F. T. Eyck, *Rapid atomic density measures for molecular shape characterization*, J. Mol. Graph. Model., vol. 19, 2001, p. 324329.
[15] J. R. Munkres, *Elements of algebraic topology*, Addison-Wesley, Redwood City, CA, USA, 1984.
[16] V. Natarajan, Y. Wang, P.-T. Bremer, V. Pascucci, and B. Hamann, *Segmenting molecular surfaces*, Comput. Aided Geom. Des. **23** (2006), no. 6, 495–509.
[17] J. J. van Wijk, *Views on visualization*, IEEE Transactions on Visualization and Computer Graphics **12** (2006), no. 4, 1000–433.