

Visual Analysis of Interactions in Multifield Scientific Data

A THESIS

SUBMITTED FOR THE DEGREE OF

Doctor of Philosophy

IN THE FACULTY OF ENGINEERING

by

Suthambhara N



Computer Science and Automation

Indian Institute of Science

BANGALORE – 560 012

NOVEMBER 2011

©Suthambhara N
NOVEMBER 2011
All rights reserved

TO

My Parents

Acknowledgements

Over the last few years, I received a lot of support from many people to make this thesis possible. In my culture, it is believed that it is impossible to thank ones family with words or even action for all the sacrifices that they make. I am thankful that I was born to such wonderful parents. The unconditional support that I received from my parents and my brother made me stronger and motivated me to do better at every step in my research.

My learning curve steeply grew after joining the research program at IISc. The most important reason for this was because of regular interactions with my research advisor, Dr. Vijay Natarajan. His guidance and critical feedback helped me to improve in all aspects of doing research; from formulating a research problem to presenting results in front of an academic audience. The work in this thesis is his as much as it is mine. I would like to thank him for his continued belief in my ability and his support for the past five years.

I am grateful to the former chairman, Prof Narasimha Murty, for his support during my initial days at IISc. I am also thankful to the chairman, Prof. Narahari for the support and encouragement given to me. There were some courses that I enjoyed a lot at IISc and the most enthralling of them was the Topics in Algorithms course taught by Dr. Kavitha Telikepalli. I thank her for those wonderful lectures. I am also grateful to Prof. Ravi Nanjundiah for his valuable comments on our experiments with the climate dataset.

I made a lot of friends after coming here and some of my friends came to study at IISc. I shall cherish the memories of my days at IISc with them. They supported me in all my endeavors and their advice on certain personal issues helped me immensely.

Throughout my studentship, I was supported by a scholarship from the Ministry of Human Resources, Govt of India. My research was also supported by funding from the Dept of Science and Technology, Govt of India (grant:SR/S3/EECE/048/2007) and travel grants from Microsoft Research and IBM, India Research Labs. I am thankful to all these organizations for their support.

I am grateful to the office staff, especially Mrs. Lalita and Mrs. Suguna for getting all my office work done so quickly and efficiently.

Publications based on this Thesis

1. Suthambhara N., Vijay Natarajan and Ravi S Nanjundiah. A Gradient-Based Comparison Measure for Visual analysis of Multifield Data. *Computer Graphics Forum (EuroVis 2011)*, 30(3), 2011, 1101-1110. Third best paper award
2. Suthambhara N. and Vijay Natarajan. Relation-aware Isosurface Extraction in Multifield Data. *IEEE Transactions on Visualization and Computer Graphics*, 17(2), 2011, 182-191
3. Suthambhara N. and Vijay Natarajan. Simplification of Jacobi sets. *In Topological Data Analysis and Visualization: Theory, Algorithms and Applications*. Valerio Pascucci, Xavier Tricoche, Hans Hagen, and Julien Tierny (editors), Springer-Verlag, Mathematics and Visualization Series, 2011, 91-102.

Presentations

- A Gradient-Based Comparison Measure for Visual analysis of Multifield Data. IEEE / Eurographics EuroVis, Bergen, 2011.
- Relation-aware Isosurface Extraction in Multifield Data. IEEE VisWeek, TVCG session, Providence, 2011.
- Simplification of Jacobi sets. TopoInVis, Snowbird, 2009.

Abstract

Data from present day scientific simulations and observations of physical processes often consist of multiple scalar fields. It is important to study the interactions between the fields to understand the underlying phenomena. A visual representation of these interactions would assist the scientist by providing quick insights into complex relationships that exist between the fields.

We describe new techniques for visual analysis of multifield scalar data where the relationships can be quantified by the gradients of the individual scalar fields and their mutual alignment. Empirically, gradients along with their mutual alignment have been shown to be a good indicator of the relationships between the different scalar variables.

The Jacobi set, defined as the set of points where the gradients are linearly dependent, describes the relationship between the gradient fields. The Jacobi set of two piecewise linear functions may contain several components indicative of noisy or a feature-rich dataset. For two dimensional domains, we pose the problem of simplification as the extraction of level sets and offset contours and describe a robust technique to simplify and create a multi-resolution representation of the Jacobi set.

Existing isosurface-based techniques for scalar data exploration like Reeb graphs, contour spectra, isosurface statistics, etc., study a scalar field in isolation. We argue that the identification of interesting isovalues in a multifield data set should necessarily be based on the interaction between the different fields. We introduce a variation density function that profiles the relationship between multiple scalar fields over isosurfaces of a given scalar field. This

profile serves as a valuable tool for multifield data exploration because it provides the user with cues to identify interesting isovalues of scalar fields.

Finally, we introduce a new multifield comparison measure to capture relationships between scalar variables. We also show that our measure is insensitive to noise in the scalar fields and to noise in their gradients. Further, it can be computed robustly and efficiently. The comparison measure can be used to identify regions of interest in the domain where interactions between the scalar fields are significant. Subsequent visualization of the data focuses on these regions of interest leading to effective visual analysis.

We demonstrate the effectiveness of our techniques by applying them to real world data from different domains like combustion studies, climate sciences and computer graphics.

Contents

Acknowledgements	i
Publications based on this Thesis	iii
Abstract	iv
1 Introduction	1
1.1 Scalar Field Visualization	3
1.2 Multifield Visualization	3
1.3 Contributions	5
1.4 Chapter Outline	8
2 Background	9
2.1 Manifolds	9
2.2 Gradients on manifolds	10
2.3 Representation of Manifolds	11
2.4 Representation of Scalar Fields	11
2.4.1 Piecewise Linear Functions	12
2.4.2 Isosurface / level set	12
3 Simplification of Jacobi Sets	13
3.1 Introduction	13
3.1.1 Motivation	13
3.1.2 Prior work and proposed approach	14
3.2 Background	14
3.2.1 Morse Theory	14
3.2.2 Reeb Graphs	15
3.2.3 Jacobi Sets	16
3.3 Simplification	16
3.3.1 Offsetting Components	18
3.3.2 Greedy Algorithm	19
3.3.3 Implementation	22
3.4 Analysis	25
3.4.1 Simplifying the input function	25

3.4.2	Effect on global comparison measure	26
3.5	Implementation for Piecewise Linear Functions	29
3.6	Applications	30
3.6.1	Visualizing Silhouettes	30
3.6.2	Combustion	31
3.7	Conclusions	33
4	Relation-aware Isosurface Extraction	34
4.1	Introduction	34
4.1.1	Related Work	35
4.1.2	Outline	37
4.2	Variation Density Function	37
4.2.1	Comparison Measure	37
4.2.2	Definition	38
4.2.3	Variation Density and Total Variation	40
4.2.4	Variation Density Profile	42
4.3	Computation	44
4.4	Applications	46
4.4.1	2D Combustion	47
4.4.2	Time Varying Combustion	49
4.4.3	Hurricane Isabel	51
4.4.4	Universe Simulation	52
4.5	Discussion	54
4.6	Conclusions and Future work	55
5	Multifield Comparison Measure	57
5.1	Introduction	57
5.1.1	Proposed Approach	58
5.1.2	Related Work	58
5.1.3	Outline	60
5.2	Multifield Comparison Measure	60
5.2.1	Matrix Norm	61
5.2.2	Comparison Measure	61
5.3	Analyzing Synthetic Functions	65
5.3.1	One / two scalar functions	67
5.3.2	Multiple / time-varying scalar functions	67
5.4	Computation	68
5.4.1	Maximum eigenvalue computation	68
5.4.2	Analysis	69
5.5	Applications	69
5.5.1	Isabel Hurricane	70
5.5.2	Global Wind Patterns	73
5.5.3	Hydrogen Combustion	77
5.6	Discussion	79

5.6.1	Limitations	80
5.6.2	Sensitivity to noise	80
5.6.3	Multifield comparison measure and PCA	81
5.7	Conclusions	81
6	Conclusions	83
	Bibliography	85
	Index	91

List of Figures

1.1	Visualization hierarchy	2
1.2	Scalar field visualization	2
1.3	Multifield visualization of a hurricane	4
2.1	Simplices of different dimensions	10
2.2	Simplicial Complex	10
3.1	Reeb graph of a manifold	15
3.2	Jacobi set of two analytic functions	16
3.3	Offsetting a level set component	17
3.4	Offset operations : merge, split, purge and create	19
3.5	Directed Reeb Graph	21
3.6	Reeb graph : Unreachable vertices	21
3.7	Simplifying the input function	26
3.8	Voronoi region of a vertex	29
3.9	Applications:Silhouettes	31
3.10	Applications:Silhouettes	32
3.11	Applications:Combustion	32
4.1	Local orthonormal coordinate system on an isosurface	38
4.2	Variation density for two functions	40
4.3	Variation density for synthetic functions	41
4.4	Computation of ψ for a two dimensional simplicial complex	44
4.5	Profiling isocontours of oxygen during a combustion simulation	47
4.6	Phases in time varying combustion	49
4.7	Hurricane Isabel	51
4.8	Universe Simulation	53
5.1	Piecewise-linear function defined on a triangle in a 2D mesh	62
5.2	Multifield comparison measure for synthetic function	66
5.3	Equivalent configurations of gradients	66
5.4	Hurricane Isabel track	70
5.5	Fronts in Hurricane Isabel at hour 10	71
5.6	Fronts in Hurricane Isabel at hour 40	71
5.7	Multifield comparison measure for wind velocities	74

5.8	Changes in storm tracks	77
5.9	Wind patterns for other climate models	78
5.10	Analyzing phases of combustion using the multifield comparison measure . .	79
5.11	Stability in the comparison measure for Isabel	80

Chapter 1

Introduction

In this thesis, we develop techniques to visualize features of interest in modern scientific data. The data is typically captured as samples over a domain or a region of interest having a geometric structure or from simulations. As an example, consider the recordings from a weather station where different quantities like wind velocity, atmospheric pressure and precipitation are collected as samples over a geographic region (domain). Each sample is a measurement of a quantity of interest. The measurements could be scalars (precipitation), vectors (wind velocity) or tensors. An elegant visual representation of these measurements can greatly help the scientist in analyzing the data. The focus of this thesis is on developing new visualization techniques for scalar scientific data.

Visualization is a process to communicate information graphically. The information to be displayed is part of data that could come from many sources like sensors, surveys, scientific computations etc. Techniques for visualization can be classified based on the type of data to be displayed. The data may have multiple forms like text, a graph or scientific measurements over a spatial region. The different areas of focus within the field of visualization classified based on the kind of data to be visualized, is shown in Figure 1.1. We refer the interested reader to the book by Ward et al. [1] for a detailed description of each topic in the figure.

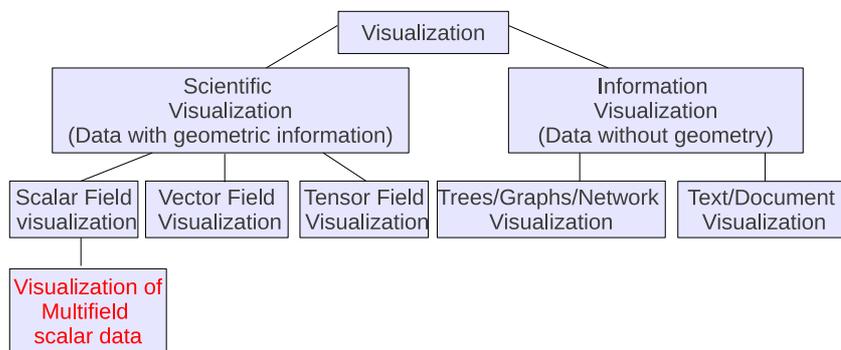


Figure 1.1: Tree depicting the classification of visualization based on the kind of information to be visualized. In this thesis, we focus on multi-field scalar field visualization.

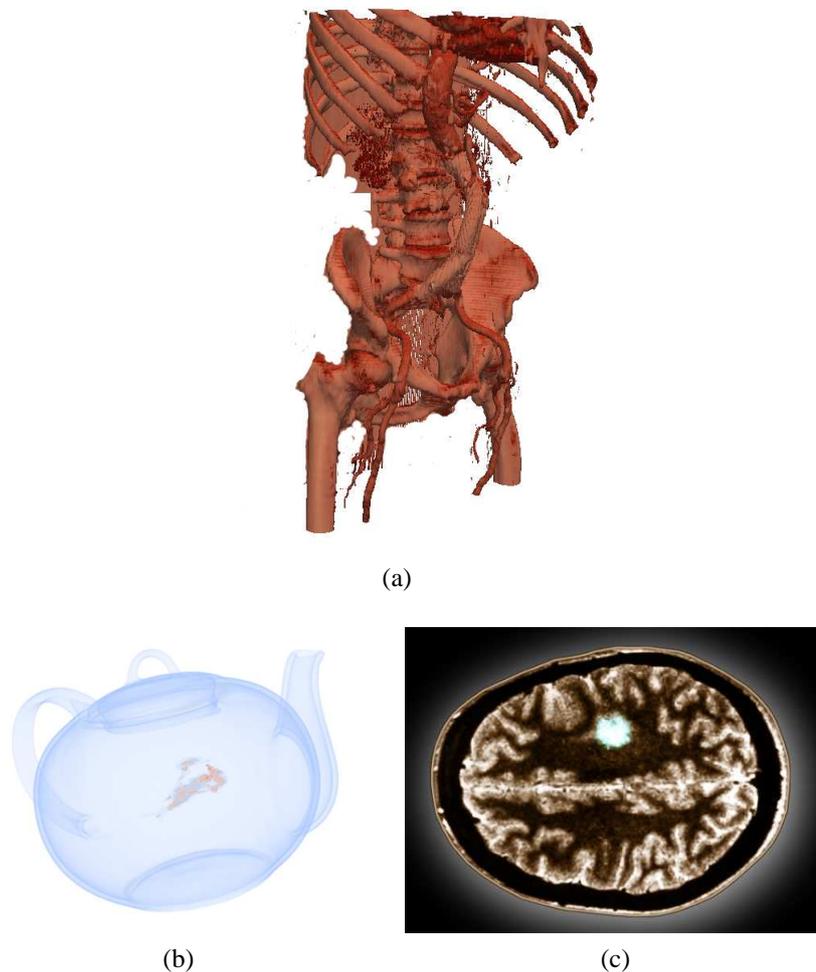


Figure 1.2: (a) Isosurface of an abdominal CT Scan (data source:<http://volvis.org>). (b) Volume rendering of a CT scan of a teapot containing a lobster inside it (data source:<http://volvis.org>). (c) MRI scan of brain shown with color coding. The bright spot is a tumor (image source:<http://www.scottcamazine.com>).

1.1 Scalar Field Visualization

A scalar function or a scalar field is a function that maps points from an n -dimensional space to real values. Scalar fields are often constructed by measuring devices or simulations. For example, during a CT scan, X-rays are passed through the body and their strength after passing is measured and a scalar field is constructed.

Visualization significantly helps in the understanding of the scalar field by allowing the user to look at the distribution of real values over the domain. The simplest way to visualize a scalar field is using colors to represent data values. Color coding gives visual representation of the scalar value at a point in relation to its neighboring values. Another useful approach is to show an isosurface/level set, which is a set of points with the same function value. Figure 1.2a shows an isosurface/level set of a CT scan of the abdomen and the pelvis. When the domain is three dimensional it is possible to view the scalar field as a single volume. A volume rendering of a CT scan of a teapot containing a lobster inside is shown in Figure 1.2b. Slices of the 3D volume displayed as 2D images also help understand the data. One such color coded slice from an MRI scan of the brain with tumor is shown in Figure 1.2c.

1.2 Multifield Visualization

Data from scientific experiments and simulations typically contain multiple scalar fields defined on a single domain. Techniques for effective visualization of single scalar fields are often inadequate for visual analysis of multifield scientific data due to the complex relationships that exist between the different scalar fields. As an example, let us consider a simulation of combustion. The data consists of scalar fields corresponding to fuel and air densities at each point on a planar domain. Effective visual representations of the scalar fields individually do not capture the interactions that exist between fuel and air. As another motivating example, we consider a simulation of a hurricane. Figure 1.3a shows the domain of simulation. Weather fronts are regions that separate air masses of different densities. It is also known that regions

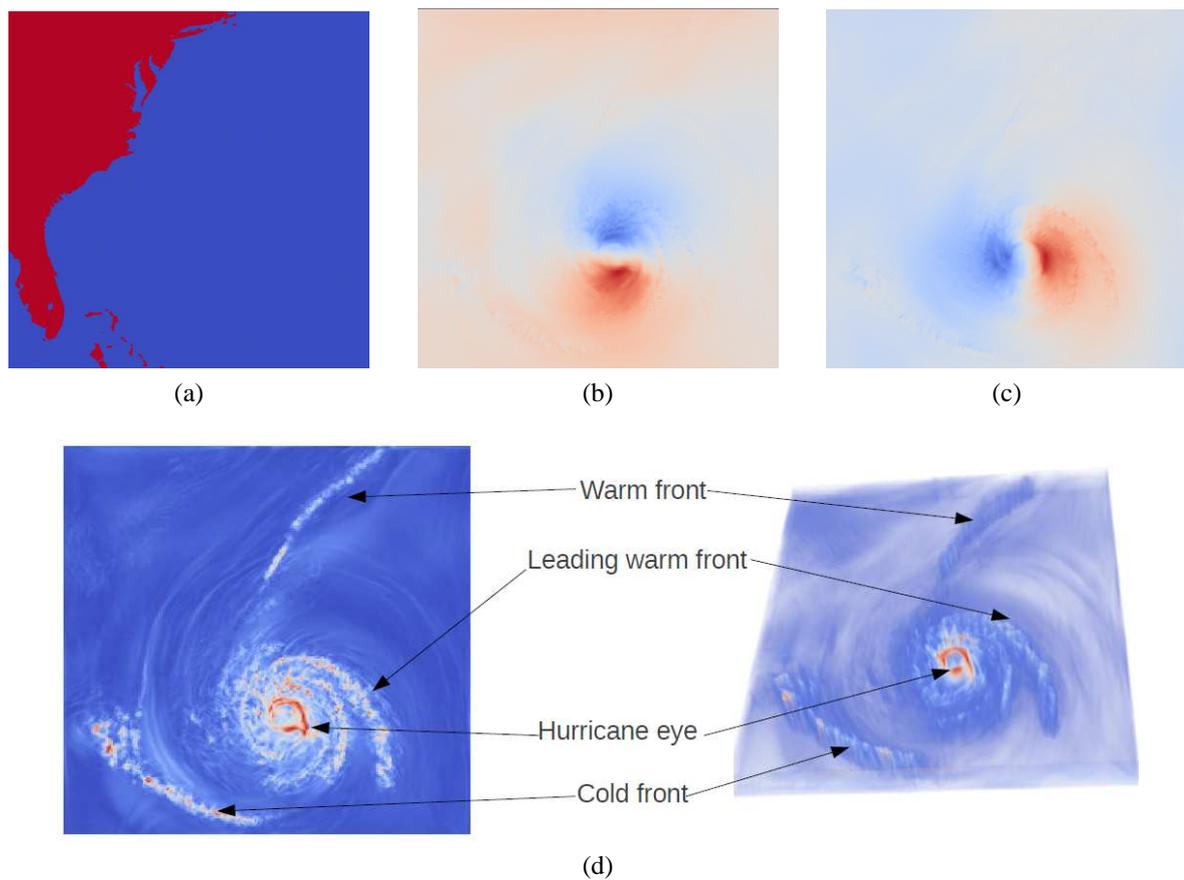


Figure 1.3: Fronts in a hurricane. (a) Region of simulation. Land mass is shown in red. (b) Volume rendering (top view) of horizontal wind speed U_f . (c) Volume rendering (top view) of horizontal wind speed V_f . (d) Volume rendering (top view) of a derived scalar field capturing rainbands at different fronts. The location of the fronts is not available from visualizations of the scalar fields U_f and V_f .

inside the front have high horizontal wind turbulence. Visualizations of the individual wind velocity components are shown in Figures 1.3b and 1.3c. We observe that the individual visualizations do not show the weather fronts. We have developed a technique (see Chapter 5) where we construct a new scalar field that captures relationships between the scalar fields. We visualize this derived field as shown in Figure 1.3d. The different weather fronts that are helpful to a climate scientist are now clearly shown in the visualization.

In all the techniques described in this thesis, we consider the lengths and mutual alignment of gradients of the individual scalar fields to quantify relationships. Gradients and their mutual alignment have been used in literature to study a wide variety of both single field and multifield data [2, 3, 4, 5, 6]. For example, the notion of critical points for a scalar fields can be extended to multiple functions by considering the Jacobi set [5].

1.3 Contributions

The goal of this thesis is to develop techniques that help in the understanding and visualization of the complex interactions that exist between multiple scalar fields in a real world data. As mentioned in the previous section, we would like to study if the gradients and their mutual alignment play a role in the relationships that exist between the different scalar variables.

We follow two approaches to visualize interactions between the different scalar fields in multifield data.

- A subset of the domain relevant to the relationships between the scalar variables under consideration is extracted. The extracted subset typically has a geometric or topological structure. The extracted subset is then visualized to gain insight into the multifield data.
- A new scalar field is derived from the scalar variables that capture interactions between the fields. A visualization of this new derived field gives an indication of the relationships that exist between the scalar variables.

In our first work, we follow the first approach to simplify Jacobi sets. Jacobi sets, introduced by Edelsbrunner et al. [5], extend the notion of critical points to multiple scalar functions. The alignment of the gradients determine the inclusion of a domain point in the Jacobi set. We interpret the Jacobi set of a pair of scalar functions as the level set of a derived function and describe a new relationship-preserving method for simplifying the Jacobi set. Specifically,

- We describe a new algorithm to compute the Jacobi set of two scalar fields using a level set interpretation. We derive a new field and extract the Jacobi set as a level set of this field. Such an interpretation helps us to naturally simplify the Jacobi set.
- We cast the simplification problem as an integer linear program. We describe an offset operation to change the topology of individual components of the Jacobi set. We then describe a greedy algorithm using offset operations to solve the linear program removing small loops. Repeated application of the simplification leads to multi resolution representation of the Jacobi set.
- We show that the change in the relationship between the functions due to simplification is upper bounded by the amount of simplification.
- We show an application of our technique to computer graphics. We compute the silhouette of a model as a Jacobi set and simplify the silhouette. We also apply our technique on a combustion data to study the relationship between fuel and air.

Multiple techniques have been proposed in the literature to identify “important” isosurfaces in scalar field data [7, 8]. However, these approaches do not have obvious extensions to multifield data. We argue that it is necessary to consider relationships between scalar fields to determine the importance of isosurfaces. As part of this thesis,

- We have defined a new function called the variation density function to measure the importance of an isosurface in multifield scientific data.

- We have described an efficient algorithm to compute a profile of the importances of iso-values, called the variation density profile, for piecewise linear functions. The algorithm is robust ensuring that the computed profile converges to the actual profile in the limit.
- We have developed a parallel implementation of the variation density profile computation using GPUs.
- We have theoretical results establishing a link between the variation density function and the well understood notion of topological persistence. This also ensures that the variation density function is insensitive to noise in the data.
- We use the variation density profile to study datasets from different domains like combustion studies, astrophysics and climate sciences. We compare our results with existing techniques and show that the variation density profile is able to identify important iso-values better.

We use the local comparison measure [3] to simplify Jacobi sets and to identify important isosurfaces using the variation density profile. The local comparison measure is defined only when the number of functions being compared does not exceed the dimension of the domain. Also, existing approaches to compare scalar functions often do not work well for more than two functions. We address these limitations by introducing a new derived scalar field called the multifold comparison measure. Specifically,

- We introduce a new multifold comparison measure to capture relationships. Visualization of the multifold comparison measure shows the interactions between different scalar variables.
- We show that the multifold comparison measure satisfies important properties like coordinate system independence and its insensitiveness to noise in the scalar fields and also their gradients.

- We describe an efficient algorithm to compute the comparison measure for piecewise linear functions.
- We apply the multifield comparison measure to study data from climate sciences and combustion studies.

1.4 Chapter Outline

In chapter 2, we describe the background required for understanding the techniques that are described in the thesis. In chapter 3, we describe a new algorithm to compute and simplify Jacobi sets. In chapter 4, we introduce the variation density function to measure relationships on an isosurface and apply it to identify important isovalues in multifield data. We describe the multifield comparison measure and its applications in chapter 5. We conclude the thesis in Chapter 6.

Chapter 2

Background

This chapter covers the necessary background. Scalar fields are defined on spatial domains with a geometric structure. In this thesis, we restrict our attention to a special class of domains called manifolds. Intuitively, a manifold is a space that resembles the Euclidean space locally, for example, a hollow sphere can be considered as a space that is locally two dimensional. The n -dimensional Euclidean space is an n -manifold.

2.1 Manifolds

A function $f : X \rightarrow Y$ between spaces X and Y is said to be a *homeomorphism* if f is bijective, continuous and the inverse of f is continuous. Spaces X and Y are said to be *homeomorphic* if there exists a homeomorphism between them. An n -manifold is a space where every point has a neighbourhood homeomorphic to \mathbb{R}^n . Each point can be represented using a local coordinate system. As in \mathbb{R}^n , it is possible to define an n -dimensional vector (called the tangent vector) at every point (technically, vectors can be defined only on a class of manifolds called differential manifolds). Recall that the length of a vector in \mathbb{R}^n is given by the square root of the inner product of the vector with itself. This notion can be extended by allowing inner products to be defined at every point on the manifold. The metric arising due to such an inner product is called the *Riemannian metric* and the manifold itself is said to be *Riemannian*. A more

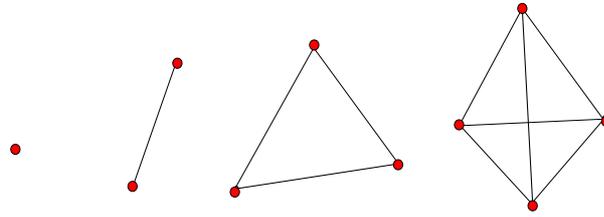


Figure 2.1: Left to right: 0-simplex (point), 1-simplex (line segment), 2-simplex (triangle) and 3-simplex (tetrahedron).



Figure 2.2: The model of a teapot represented as a simplicial complex. The elements of the simplicial complex are vertices (0-simplices), edges (1-simplices) and faces (2-simplices). The simplicial complex is also a 2-manifold.

rigorous and technical explanation can be found in the book by Guillemin and Golubitsky [9].

We consider only Riemannian manifolds throughout this thesis.

2.2 Gradients on manifolds

The gradient of a smooth function defined on \mathbb{R}^n at a point $x \in \mathbb{R}^n$ is the vector whose components are the partial derivatives of the function along each coordinate axis. The direction of the gradient vector signifies the direction of steepest ascent of the function. The maximum rate of change is given by the length of the gradient. If f is a smooth function defined on a Riemannian manifold and (x_1, x_2, \dots, x_n) is a local coordinate system such that the unit tangent vectors denoted by $(\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n})$ form an orthonormal basis. The gradient of f at x is defined as the tangent vector $\nabla f(x) = (\frac{\partial f}{\partial x_1}(x), \dots, \frac{\partial f}{\partial x_n}(x))$. If the gradient vanishes at x , then x is said to be a *critical point* of f .

2.3 Representation of Manifolds

A Manifold can be represented in multiple ways on a computer. For example, it can be represented as a collection of cells that do not intersect with one another. A *structured grid* is a representation of a manifold with structured geometry where the cells are quadrilaterals (2D) or cuboids (3D). The cells are attached to each other sharing edges and vertices and partition the domain. *Rectilinear grids* are structured grids where all cells are rectangles (2D) or rectangular cuboids (3D). The geometry of the manifold could be irregular where the locations of points on the manifold are specified arbitrarily.

Manifolds with irregular geometry can be represented using simplicial complexes. For $0 \leq k \leq n$, a *k-simplex* in \mathbb{R}^n is the convex hull of $k + 1$ affinely independent points. The interior of a simplex is the set of points in the simplex that do not lie on the boundary. Define the interior of a 0-simplex to be itself (see Figure 2.1). A *simplicial complex* K is a collection of simplices such that all faces of simplices in the simplicial complex also belong to the complex and the intersection of any two simplices is empty or a face common to both. Figure 2.2 shows a simplicial complex representing a 2-manifold. The *dimension* of K is the dimension of the simplex in K with the highest dimension.

2.4 Representation of Scalar Fields

Scalar data is often available as values defined over a collection of points on a manifold. For visual analysis, it is often necessary to get a continuous representation of the data. This process of reconstructing the data into a continuous representation can be accomplished using different kinds of interpolation like linear, bilinear and trilinear interpolation [10]. Applications requiring higher order continuity in the scalar field use techniques like cubic interpolation [11].

2.4.1 Piecewise Linear Functions

In this thesis, whenever the scalar field is defined on a simplicial complex, we use piecewise linear interpolation to get a continuous representation of the field. Let K be a simplicial complex and $\bar{f} : \text{vertices}(K) \rightarrow \mathbb{R}$ be a real valued function defined on the vertices of K . We construct a *piecewise-linear function* $f_p : K \rightarrow \mathbb{R}$ by linearly extending \bar{f} within each simplex. The position of any point x in the interior of a k -simplex can be written uniquely as a convex sum of the positions of vertices of the simplex, i.e, $x = \sum_{i=1}^{k+1} \gamma_i v_i$ with $\sum_{i=1}^{k+1} \gamma_i = 1$. The tuple $(\gamma_1, \dots, \gamma_{k+1})$ is called the *barycentric* coordinate of x . Define $f_p(x) = \sum_{i=1}^{k+1} \gamma_i \bar{f}(v_i)$. The function f_p is continuous, linear within each simplex, and agrees with \bar{f} at the vertices of K . The gradient of f_p is well defined in the interior of a simplex and is a constant vector because f_p is linear within the simplex.

2.4.2 Isosurface / level set

Let f be a smooth real valued function defined on an n -manifold. The *level set* at $c \in \mathbb{R}$ is defined as the preimage $f^{-1}(c)$. The level set at c is always an $n - 1$ manifold if the gradient of f does not vanish anywhere on the level set. Level sets are also called *isocontours* or *isolines* for two dimensional manifolds and *isosurfaces* for three dimensional manifolds. When the scalar field is defined on a rectangular grid, an isosurface can be extracted efficiently using the marching cubes algorithm [12]. A similar algorithm called marching tetrahedra [13] can be used for extracting isosurfaces of piecewise-linear functions defined on a 3 dimensional simplicial complex.

Chapter 3

Simplification of Jacobi Sets

In multifield scalar data, the linear dependence between the gradients of the scalar fields can be used to study relationships between them. The Jacobi set is a subset of the domain that captures this linear dependence. In this chapter, We present a new algorithm to compute and simplify the Jacobi set.

3.1 Introduction

3.1.1 Motivation

The Jacobi set extends the notion of critical points to multiple functions and helps describe the relationship between multiple scalar functions. Edelsbrunner et al. [14] have shown that the Jacobi sets can be used to compute a comparison measure between two scalar functions. Bennett et al. [15] have used the Jacobi set to represent tunnels and the silhouette of a mesh, both of which are subsequently used to compute a cross parameterization. Jacobi sets have also been used to track features of time-varying events such as molecular interactions and combustion simulation [16]. All the above applications face a common challenge, namely the presence of degenerate regions and noise in the data. The number of components of the Jacobi set is often more than what can be visually comprehended. So, it is necessary to simplify the

Jacobi set. The simplification can be accomplished either using the notion of persistence [17], or otherwise.

3.1.2 Prior work and proposed approach

In their paper, Bremer et al. [16] have described a method to remove noise in the Jacobi set for time varying data. The persistence of a component of the Jacobi set is the time interval between its birth and death. This measure has been used to remove components that are either noise in the data or unimportant features. Extending this for general functions is nontrivial and hence a more complete approach with guaranteed error bounds is required. We pose the problem of computing Jacobi sets as the computation of a level set of a function defined on the input manifold. Jacobi set simplification is accomplished by simplifying the level set. We also ensure that the change in relationship between the functions due to simplification does not exceed a given input threshold.

3.2 Background

The simplification algorithm makes use of the Reeb graph to reduce the components in the Jacobi set. In this section, we give a brief introduction to Morse theory, Reeb graphs and Jacobi sets.

3.2.1 Morse Theory

Morse theory studies the relationship between functions and domains. Let \mathbb{M} be a smooth Riemannian 2-manifold. Let f be a smooth function defined on \mathbb{M} and (x_1, x_2) be a local coordinate system such that the unit tangent vectors $(\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2})$ form an orthonormal basis with respect to a Riemannian metric. The gradient of f at x is defined as the vector $\nabla f(x) = (\frac{\partial f}{\partial x_1}(x), \frac{\partial f}{\partial x_2}(x))$.

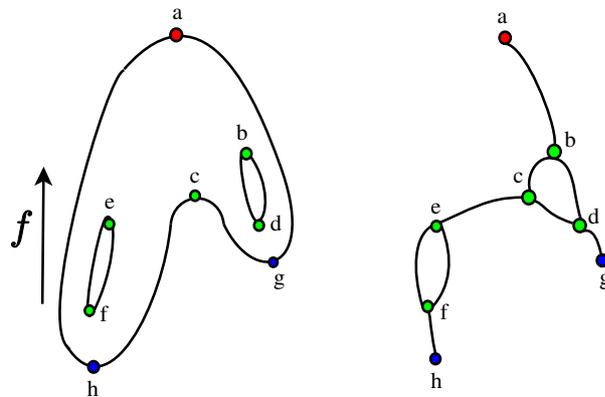


Figure 3.1: **Left:** A two-holed 2-manifold and the height function defined on it. Points in blue, green, and red correspond to minima, saddle, and maxima of the function, respectively. **Right:** The Reeb graph of the height function. Loops in the Reeb graph correspond to holes in the manifold.

A point x is a critical point of f if $\nabla f(x)$ is the zero vector. The function f is called a *Morse function* if the Hessian

$$\mathcal{H}_f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \frac{\partial^2 f}{\partial x_2 \partial x_1}(x) \\ \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \frac{\partial^2 f}{\partial x_2^2}(x) \end{bmatrix}$$

is non-singular at all critical points. Critical points are classified based on the eigenvalues of the Hessian. A critical point is a *minimum* if both the eigenvalues are positive. A *maximum* is a critical point with both eigenvalues negative and a *saddle* has one positive and one negative eigenvalue.

3.2.2 Reeb Graphs

The *Reeb graph* of f is obtained by contracting connected level set components to points. Nodes in a Reeb graph correspond to critical points of f , see Figure 3.1. For $c \in \mathbb{R}$, a level set at c is the preimage $f^{-1}(c)$ (see Section 2.4.2). The level sets *sweep* the domain as we increase c over the range of the function f . During a sweep over the domain, the topology of the level set changes at critical points of f . If the sweep is in the direction of increasing

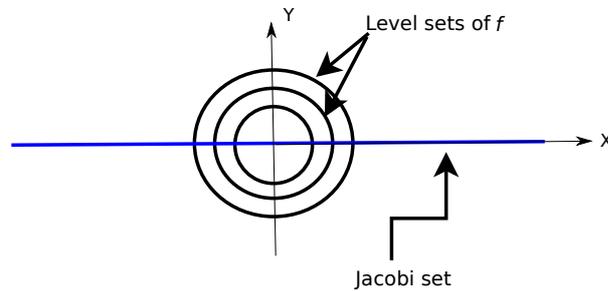


Figure 3.2: Jacobi set of analytic functions $f = x^2 + y^2$ and $g = x$ defined on \mathbb{R}^2 . The gradients of f and g align on the Jacobi set.

function value, level set components are created at minima, they merge or split at saddles, and are destroyed at maxima. Given a sweep direction, saddles may be classified as split or merge saddles depending on the change in the topology of level sets at these points.

3.2.3 Jacobi Sets

The *Jacobi set* of two Morse functions f and g defined on a 2-manifold \mathbb{M} is the collection of points where the gradients of the functions align with each other or one of the gradients vanish (see Figure 3.2). Alternately, the Jacobi set can be described as the collection of critical points of the family of functions $f + \lambda g$, $\lambda \in \mathbb{R}$:

$$\mathbb{J} = \{x \in \mathbb{M} \mid x \text{ is a critical point of } f + \lambda g \text{ or of } \lambda f + g\}$$

Note that the Jacobi set contains critical points of f and g . Edelsbrunner and Harer [3] used this alternate description to compute Jacobi sets of piecewise linear functions. They also showed that the Jacobi set of two Morse functions is a smoothly embedded 1-manifold in \mathbb{M} .

3.3 Simplification

We prefer to use the description of the Jacobi set as the level set of a gradient-based comparison measure [14] because it leads us to a natural algorithm for computing Jacobi sets. Let \mathbb{M} be a

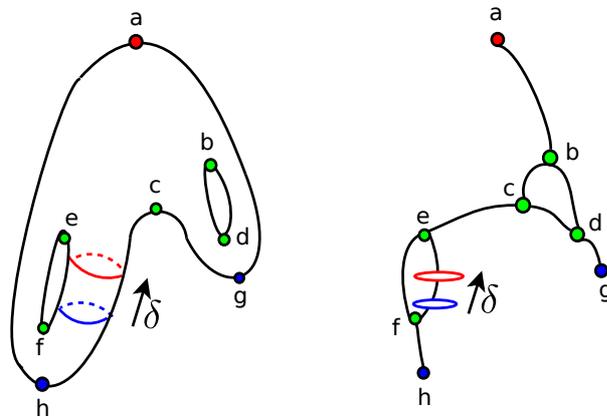


Figure 3.3: Offsetting a level set component. **Left:** Level set components on the manifold. **Right:** Offsetting a level set component (blue) to another component (red) along an edge of the Reeb graph.

2-manifold smoothly embedded in \mathbb{R}^3 . The *local comparison measure*, κ_x , at a point $x \in \mathbb{M}$ for two Morse functions f and g is defined as $\kappa_x = \|\nabla f(x) \times \nabla g(x)\|$. Assuming \mathbb{M} is orientable, we define the *sign extended comparison measure*, κ_x^S , at the point x with unit normal \hat{n} as $\kappa_x^S(f, g) = (\nabla f \times \nabla g) \cdot \hat{n}$. The sign extended comparison measure is a function defined on the manifold \mathbb{M} and the Jacobi set can be described as the set of points where κ_x^S equals zero, i.e. the zero level set of κ_x^S , $\mathbb{J} = \kappa_x^{-1}(0) = \kappa_x^{S-1}(0)$.

The Jacobi set often contains spurious loops because of noise and degeneracies in the data. Simplification of the Jacobi set refers to the reduction in number of components of \mathbb{J} with minimal change to the relationship between the two input functions.

The relationship between the functions is quantified by the *global comparison measure* κ , which is equal to the comparison measure integrated over the manifold and normalized by the total area [14].

$$\kappa = \frac{1}{\text{Area}(\mathbb{M})} \int_{x \in \mathbb{M}} \kappa_x dA_x,$$

where dA_x is the area element at x .

3.3.1 Offsetting Components

The Jacobi set components are altered by computing offset level set components. Let p and p' be two level set components such that their corresponding points on the Reeb graph are connected by a monotone path (path with monotonically increasing or decreasing function values). The level set component p is said to be offset to p' if it is replaced by the component p' . The cost of an offset operation is equal to the hypervolume of the swept region, which is computed as an integral over the swept region R of the domain:

$$H = \frac{1}{\text{Area}(\mathbb{M})} \int_{x \in R} \kappa_x dA_x. \quad (3.1)$$

Figure 3.3 shows a level set component offset upwards by a hypervolume δ . The direction of offset is upward if the function value increases and downward otherwise. We simplify the Jacobi set by computing offsets in an appropriate direction.

The following basic offset operations are used in the simplification process.

Merge. Two components whose edges share a common saddle are offset to the saddle so that they merge. The merged component is further offset by a small value resulting in a single component.

Split. A component is offset to a saddle and is further offset by a small value resulting in a split.

Purge. A component is offset to a local maximum or minimum. A further offset by a small value removes the component.

Create. A component is created at a local maximum or minimum and offset by a small value.

Figure 3.4 illustrates the basic offset operations, using the Reeb graph. The Reeb graph is naturally suited to represent the offsets because it traces the connected components of the level sets. Only two operations result in a reduction in the number of components. Temporary

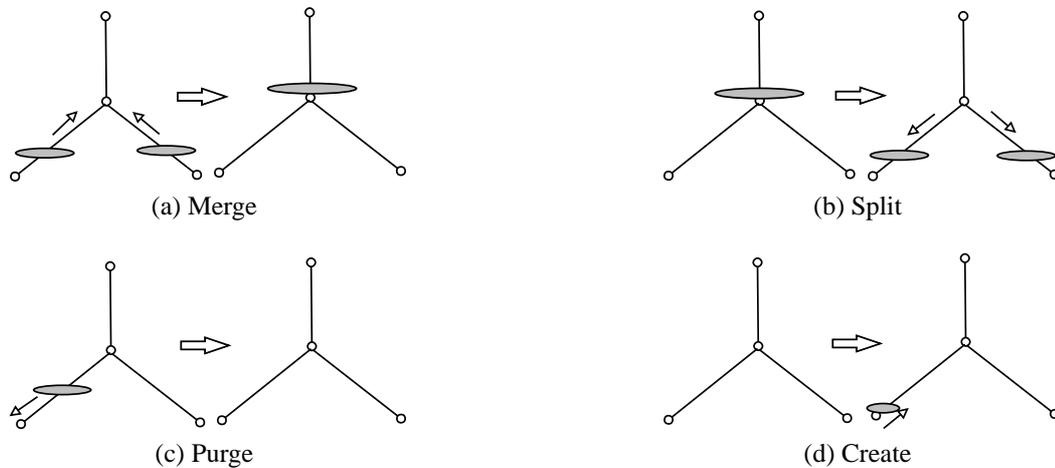


Figure 3.4: Different offset operations used during simplification. All offsets are shown against the Reeb graph of κ_x^S .

splits may be required to obtain a small number of components. We ensure that the number of splitting operations is lower than the number of component merging operations. We show in the next section that twice the total hypervolume swept during the operations is an upper bound over the total change in relationship between the functions.

The first step in the simplification procedure is the computation of the Reeb graph for κ_x^S . Arcs in the Reeb graph that contain the zero level set are also identified.

3.3.2 Greedy Algorithm

The required simplification is specified as a percentage of the global comparison measure. The corresponding hypervolume threshold, i.e., the total hypervolume allowed for the operations is calculated next. Since each simplification operation involves exactly one critical point, we can represent an offset by a critical point. We first augment the Reeb graph by inserting dummy nodes at level zero. This augmented graph is transformed into a directed graph by replacing each arc uv with a directed arc uv (arc from u towards v), if $|\kappa_v^S| > |\kappa_u^S|$, see Figure 3.5 . Each

node is then assigned a profit $P(v)$ given by

$$P(v) = \begin{cases} 1 & \text{if } v \text{ is a dummy node} \\ in(v) - out(v) & \text{otherwise,} \end{cases}$$

where $in(v)$ and $out(v)$ represent the indegree and outdegree of v in the directed graph. The profit for a non dummy node signifies the reduction in number of Jacobi set components if the operation corresponding to the node is chosen. The optimal simplification can now be formulated as an integer linear program (ILP) that maximizes profit. The variables in the ILP correspond to nodes of the directed Reeb graph.

$$\max \sum P(v)x_v$$

subject to constraints

$$\sum C(v)x_v \leq T$$

$$x_v - x_u \leq 0 \text{ for a directed arc } uv$$

$$x_u + x_v \leq 1 \text{ } u, v \text{ adjacent to a common dummy node}$$

$$x_u, x_v \in \{0, 1\}$$

The cost $C(v)$ for each simplification operation is the sum of hypervolumes of the incoming arcs. T is the threshold given as input. A simplification operation is performed on a node if the corresponding variable in the ILP is set to one. The first constraint bounds the total hypervolume for the simplification. The second constraint enforces a dependency between variables corresponding to a directed arc uv . This dependency captures the fact that a simplification operation at v can be performed only after a level set component has been offset through the node u . At dummy nodes, there is a choice to perform an offset in either of the directions but not both. This choice is modeled in the third constraint. The ILP is a variant of the knapsack

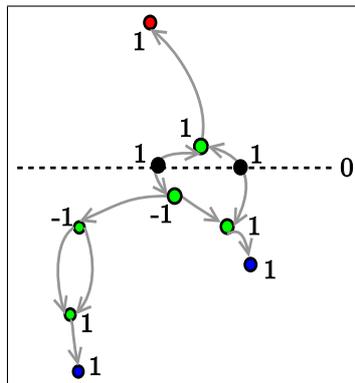


Figure 3.5: Directed Reeb graph. The dotted line in the figure shows level 0. The dummy vertices are shown in black on the zero line. The profit for each node is also shown.

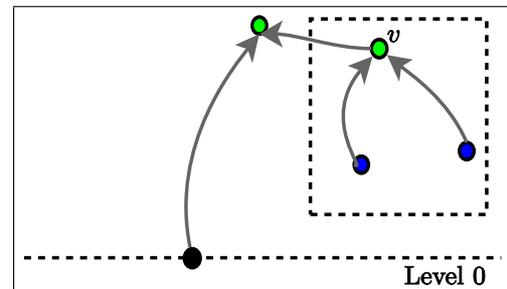


Figure 3.6: A Section of a Reeb graph with unreachable vertices shown in the boxed rectangle. The unreachable component prevents the algorithm to proceed beyond the merge saddle v .

problem with dependencies among objects. Though a solution to the above ILP corresponds to the optimal simplification, the computation is slow in practice. So, we resort to a greedy strategy that chooses the least cost offset operation at every step until the threshold is reached. The greedy strategy has an additional advantage; it enables the creation of a multi-resolution representation of the Jacobi set.

The greedy algorithm requires all nodes to be stored in a priority queue. The priority queue is initialized with all possible simplification operations and updated with new operations that may become valid after an offset is performed. We define a node of the directed Reeb graph as *unreachable* if it cannot be reached by a path from a dummy node and *reachable* otherwise. Unreachable nodes may become obstacles that prevent offset operations. For example, a saddle with an incoming arc from an unreachable node prevents a merge operation, see Figure 3.6. Let G denote the directed Reeb graph and H denote the subgraph of G containing all unreachable vertices. A component J of H is a connected component in the undirected version of H . The cost of removing J is the sum of the cost of all edges of G that have at least one end point in J . If the algorithm is not able to proceed due to some obstacles, then least cost components of unreachable vertices are removed from G until a valid operation is identified. Finally, we

extract offset components using seed sets stored in the Reeb graph [8]. We also ensure that the number of simplification operations with negative profits is smaller than a constant fraction of the operations with positive profits. This ensures that the number of components decreases as a result of simplification.

Procedure Initialize

Input: Directed Reeb Graph $G(V, E)$.

Returns: Initialized list L

```

1: mark all edges  $uv \in E$  as 'NONE'.
2: List  $L \leftarrow \{\emptyset\}$ 
3: for each dummy node  $u$  do
4:   for each outgoing edge  $uv$  from  $u$  do
5:     mark  $uv$  'READY'
6:     if all incoming edges of  $v$  are marked 'READY' then
7:       add  $v$  to  $L$ .
8:     end if
9:   end for
10: end for
11: return  $L$ 

```

3.3.3 Implementation

The idea of the greedy approach is to schedule the best possible offset operation that can be performed. This would require all vertices to be stored in a priority queue and updated with new operations that may unravel after an operation is performed. The INITIALIZE procedure populates a list with all possible initial simplification operations. Marking an edge as READY signifies a component's readiness to be offset.

The SIMPLIFYJACOBISET procedure greedily chooses the least cost operation and executes it. Marking an incoming edge as DONE signifies an offsetting along the edge. We define a node of the directed Reeb graph G as *unreachable* if it cannot be reached by a path from a dummy node and *reachable* otherwise. A saddle with an incoming edge from an unreachable node prevents a merge. If we denote H as the subgraph of G containing all unreachable vertices, a component of H is a connected component in the undirected version of H . The cost

Procedure Extract

Input: Directed Reeb graph $G(V, E)$ with marked edges from Procedure SIMPLIFYJACOBISET().

Returns: List X of components of the simplified Jacobi set.

- 1: List $K \leftarrow \{\phi\}$
- 2: **for** each node $v \in V$ **do**
- 3: **if** $in(v) = 0$ or all incoming edges of v are marked 'DONE' **then**
- 4: **if** $out(v) = 0$ or all outgoing edges from v are marked 'READY' **then**
- 5: add v to K .
- 6: **end if**
- 7: **end if**
- 8: **end for**
- 9: **for** each node v in K **do**
- 10: **if** v is a dummy node **then**
- 11: add component corresponding to v in the Reeb graph to X .
- 12: **else**
- 13: **for** each outgoing edge vw **do**
- 14: add component corresponding to a small constant offset along vw to X
- 15: **end for**
- 16: **end if**
- 17: **end for**
- 18: **return** X

Procedure SimplifyGraph

Input: Directed Reeb graph $G(V, E)$, Sub graph H of G containing unreachable vertices.

Returns: Component J after removing it from G

- 1: $J \leftarrow$ least cost component from H .
- 2: **if** $J \neq \phi$ **then**
- 3: remove all vertices in J along with their edges from G
- 4: **end if**
- 5: **return** J

Procedure SimplifyJacobiSet

Input: Directed Reeb graph $G(V, E)$, List L from Procedure INITIALIZE() and Threshold T .

Returns: None

```

1: while  $T \geq 0$  do
2:   while  $empty(L)$  and  $T \geq 0$  do
3:      $J \leftarrow SimplifyGraph$ 
4:     if  $J = \phi$  then
5:       goto line 35
6:     end if
7:      $T = T - cost(J)$ 
8:     for each edge  $uv$  with  $u \in J, v \in V - J$  do
9:       mark  $uv$  'READY'.
10:      if all incoming edges of  $v$  are marked 'READY' then
11:        add  $v$  to  $L$ 
12:      end if
13:    end for
14:  end while
15:  Choose  $v$  from  $L$  with least  $C(v)$  and  $C(v) \leq T$ 
16:  if no such  $v$  exists then
17:    goto line 35
18:  end if
19:  for each incoming edge  $uv$  do
20:    if  $u$  is a dummy node then
21:      mark all outgoing edges  $uw, v \neq w$  'NONE'.
22:      remove  $w$  from  $L$ 
23:    end if
24:    mark  $uv$  'DONE'
25:  end for
26:  for each outgoing edge  $vw$  do
27:    mark  $vw$  'READY'
28:    if all incoming edges of  $w$  are marked 'READY' then
29:      add  $w$  to  $L$ 
30:    end if
31:  end for
32:  remove  $v$  from  $L$ 
33:   $T = T - C(v)$ 
34: end while
35: return

```

of a component J denoted by $cost(J)$ is the sum of the cost of all edges of G which have at least one end point in J . The SIMPLIFYGRAPH removes a least cost component of unreachable vertices from G so that the simplification can proceed.

The final step extracts the simplified Jacobi set. Vertices in G that have all incoming edges marked as DONE and outgoing edges marked READY represent the final offset components after simplification.

3.4 Analysis

In this section we show that twice the hypervolume swept during a simplification operation is an upper bound over the change in the relationship between the input functions.

3.4.1 Simplifying the input function

We do not change the function values in our experiments. However, we now compute changes to the function f caused by a small offset in order to obtain the upper bound. Figures 3.7a and 3.7b depict the changes to the function f after offsets in the up and down directions respectively. An upward offset introduces critical points at E and F of f restricted to level set I of g . To accomplish this, the function values at E and F can be interchanged to become $f(F)$ and $f(E)$ respectively. Within level set II of g , the critical points of f move from B and C to A and D respectively. The function f restricted to level set II between A and D is made monotone to achieve this movement of critical points. The function values at A and D do not change and therefore the new pair have a reduced persistence. Downward offset destroys the critical point pair E and F and the restricted function f between E and F is made monotone. The function values at E and F are interchanged to become $f(F)$ and $f(E)$ respectively. Within level set II of g , critical points move from A and D to B and C respectively.

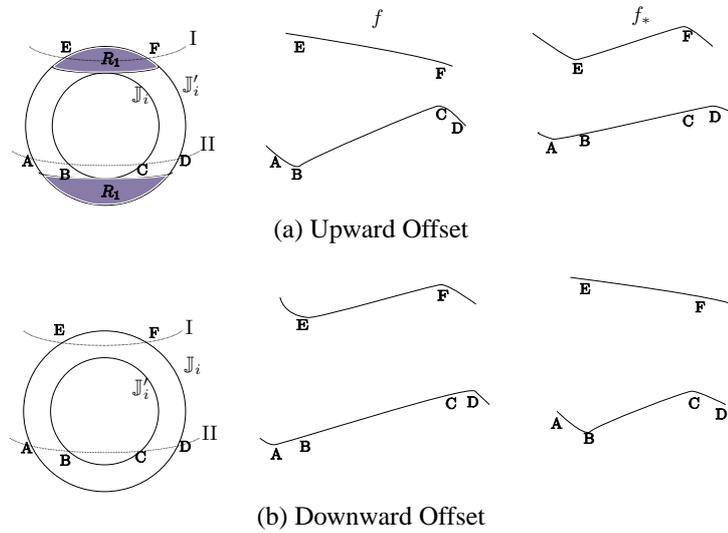


Figure 3.7: Simplifying the input function. The left column shows a Jacobi set component \mathbb{J}_i and its offset version \mathbb{J}'_i . The dashed lines are level sets of the function g . The center column shows f restricted to the level sets I and II. The right column shows the simplified function f_* that corresponds to the offset Jacobi set component \mathbb{J}'_i .

3.4.2 Effect on global comparison measure

As shown by Edelsbrunner et al. [14], the global comparison measure can be computed by considering restrictions of one function on the isocontours of the second function. The global comparison measure in this form is given by

$$\kappa = \frac{2}{\text{Area}(\mathbb{M})} \int_{v \in \mathbb{J}} \text{sign}(v) f(v) dg,$$

where $\text{sign}(v)$ is defined as

$$\text{sign}(v) = \begin{cases} +1 & \text{if } v \text{ is a maximum of } f|_{g^{-1}(g(v))} \\ -1 & \text{otherwise.} \end{cases}$$

Let \mathbb{J}_i denote the i^{th} component of the Jacobi set. Define

$$\kappa_i = \frac{2}{\text{Area}(\mathbb{M})} \int_{v \in \mathbb{J}_i} \text{sign}(v) f(v) dg.$$

κ_i can be interpreted as the contribution of \mathbb{J}_i to the global comparison measure, $\kappa = \sum_i \kappa_i$.

Since the change to the function f corresponding to an offset is local to the region of the component, we will now compute the change in κ_i corresponding to an upward offset. If f_* is the modified function, the change in κ_i is given by

$$|\delta \kappa_i| = \frac{2}{\text{Area}(\mathbb{M})} \left| \int_{v \in \mathbb{J}'_i} \text{sign}(v) f_*(v) dg - \int_{v \in \mathbb{J}_i} \text{sign}(v) f(v) dg \right|.$$

Let R be the region of \mathbb{M} swept during the offset and R_1 be the region where the level sets of g do not intersect \mathbb{J}_i (shaded region in Figure 3.7a). The integral over \mathbb{J}'_i can be rewritten as a sum of integrals over two regions:

$$\begin{aligned} |\delta \kappa_i| = \frac{2}{\text{Area}(\mathbb{M})} & \left| \int_{v \in \mathbb{J}'_i \cap R_1} \text{sign}(v) f_*(v) dg - \int_{v \in \mathbb{J}_i} \text{sign}(v) f(v) dg \right. \\ & \left. + \int_{v \in \mathbb{J}'_i \cap (R - R_1)} \text{sign}(v) f_*(v) dg \right|. \end{aligned} \quad (3.2)$$

Consider the level sets I in Figure 3.7a. The difference between function values at E and F can be written as

$$f_*(F) - f_*(E) = f(E) - f(F) = \int_F^E \|\nabla f_t(x)\| dl.$$

Here, $\nabla f_t(x)$ represents the tangential component of $\nabla f(x)$ along the level sets and dl is the length element along the level set. The integral of $\text{sign}(v) f_*(v)$ over $\mathbb{J}'_i \cap R_1$ can be rewritten as

an integral over R_1 using the above expression,

$$\int_{v \in \mathbb{J}'_f \cap R_1} \text{sign}(v) f_*(v) dg = \iint_{x \in R_1} \|\nabla f_t(x)\| dl dg.$$

Let du be the length element orthogonal to the level set. The area element is given by $dl du$.

Using the fact that $dg = \|\nabla g(x)\| du$,

$$\begin{aligned} \int_{v \in \mathbb{J}'_f \cap R_1} \text{sign}(v) f_*(v) dg &= \iint_{x \in R_1} \|\nabla f_t(x)\| \|\nabla g(x)\| dl du = \int_{x \in R_1} \|\nabla f(x) \times \nabla g(x)\| dA_x \\ &= \int_{x \in R_1} \kappa_x dA_x. \end{aligned} \quad (3.3)$$

Consider the level set Π of g in Figure 3.7a:

$$f_*(A) = f(A) = f(B) + \int_B^A \|\nabla f_t(x)\| dl$$

and

$$f_*(D) = f(D) = f(C) - \int_D^C \|\nabla f_t(x)\| dl.$$

Combining the above two equations,

$$\begin{aligned} (f(C) - f(B)) - (f_*(D) - f_*(A)) &= (f_*(A) - f(B)) + (f(C) - f_*(D)) \\ &= \int_B^A \|\nabla f_t(x)\| dl + \int_D^C \|\nabla f_t(x)\| dl. \end{aligned}$$

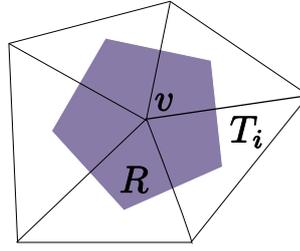


Figure 3.8: Consider the vertex v and its adjacent vertices as a point set. The neighborhood R of a vertex v on a piecewise linear surface is represented by the Voronoi region of v .

All pairs of points $A, D \in \mathbb{J}'_i \cap (R - R_1)$ have a corresponding pair $B, C \in \mathbb{J}_i$. So, we have

$$\begin{aligned}
 \left| \int_{v \in \mathbb{J}'_i \cap (R - R_1)} \text{sign}(v) f_*(v) dg - \int_{v \in \mathbb{J}_i} \text{sign}(v) f(v) dg \right| \\
 &= \iint_{x \in (R - R_1)} \|\nabla f_i(x)\| \|\nabla g(x)\| dl du \\
 &= \int_{x \in (R - R_1)} \kappa_x dA_x. \tag{3.4}
 \end{aligned}$$

Substituting (3.4) and (3.3) in (3.2) and using the triangle inequality,

$$|\delta k_i| \leq \frac{2}{\text{Area}(\mathbb{M})} \int_R \kappa_x dA_x = 2H.$$

The above inequality can be similarly derived for the downward offset. Thus, the hypervolume is a conservative estimate of the change in relationship between f and g caused by an offset.

3.5 Implementation for Piecewise Linear Functions

Scalar scientific data is often represented by piecewise linear functions on triangle meshes, where the gradient and hence κ_x^S is not defined at vertices of the mesh. Given a vertex v of the triangle mesh, its neighborhood is the Voronoi region as shown in Figure 3.8. Meyer et al. [18] used the Voronoi region to define discrete differential operators with minimal numerical error for triangulated surfaces. Let T_1, T_2, \dots, T_i be triangles that intersect the neighborhood R of v .

The sign extended comparison measure κ^S is constant within each of the regions $T_i \cap R$. We follow Meyer et al. to define κ_v^S as the average value of the sign extended measure over R ;

$$\kappa_v^S = \frac{1}{Area(R)} \sum_{i=1}^t \kappa_i^S Area(T_i \cap R),$$

where κ_i^S is the value of the sign extended comparison measure at a point that lies in the interior of T_i . Note that the gradients of f and g are constant in the interior of a triangle and hence κ_x^S is also constant within a triangle. The sign extended comparison measure is stored at vertices and a linear approximation is used within the edges and triangles. This approximation does not introduce significant artifacts in practice. The zero level set can be extracted using a marching triangles algorithm or from seed sets computed using a Reeb graph of κ_x^S .

3.6 Applications

We demonstrate the usefulness of the simplified Jacobi set using two different applications. Our approach to the definition and simplification of Jacobi sets is particularly useful when studying the relationship between two functions using their gradients.

3.6.1 Visualizing Silhouettes

Given a view direction d in \mathbb{R}^3 and a 2-manifold \mathbb{M} embedded smoothly in \mathbb{R}^3 , the silhouette is the set of points in \mathbb{M} where the tangent plane is parallel to d . Consider a Cartesian coordinate system with the z -axis along the view direction d . The Jacobi set of the two scalar fields $f(x, y, z) = x$ and $g(x, y, z) = y$ is the required silhouette. The silhouette of a model of the hand¹ is shown in Figure 3.9c. The model is shown in the original orientation in Figure 3.9a. The view direction is perpendicular to the plane of paper. The orientation of the model has been changed for a better view of the computed silhouette in Figures 3.9b and 3.9c. As seen from the

¹The models of hand and torso were downloaded from the AIM@SHAPE repository (<http://www.aimatshape.net/>).

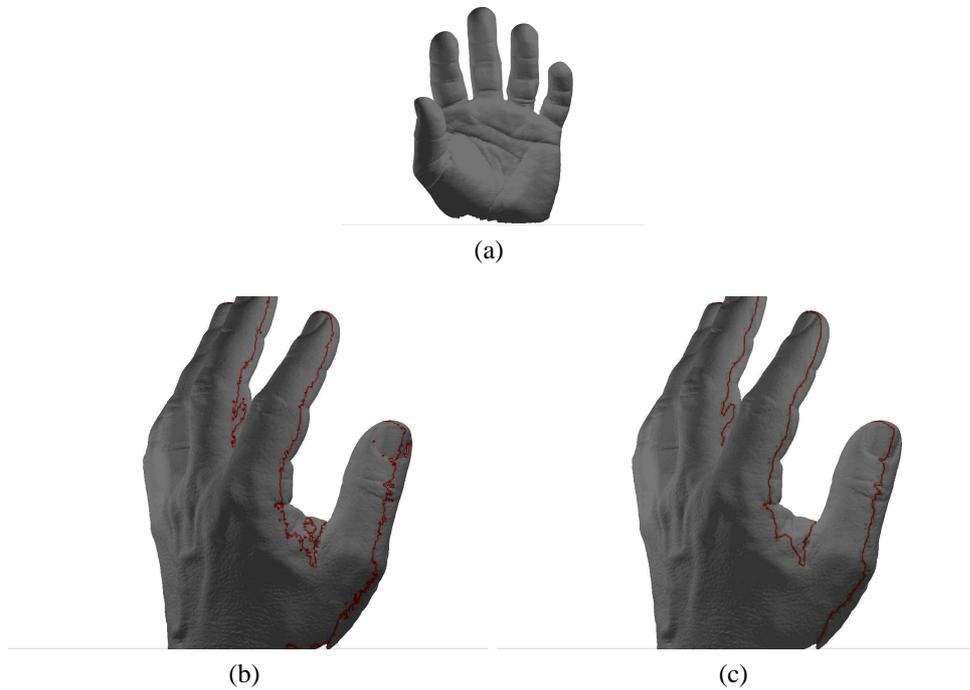


Figure 3.9: Silhouettes. (a) Model of a hand in its original orientation. (b) Silhouette when viewed from a different angle. (c) Simplified silhouette.

figure, the silhouette has many components that are unimportant caused by small bumps in the model and the silhouette itself appears to contain noise. The simplification process removes small components because their removal does not adversely affect the relationship between the fields f and g used to compute the silhouette. We found that simplification using 2% threshold removed all noise. The Jacobi set was simplified using the greedy algorithm. Similar results are shown for another model in Figure 3.10

3.6.2 Combustion

We apply our algorithm to study a time varying dataset from the simulation of a combustion process.² This application demonstrates the use of simplification when handling degenerate data. Degeneracies occur when κ_x is zero within a region, resulting in the Jacobi set containing

²We would like to thank Jackie Chen and Valerio Pascucci for providing the combustion data .

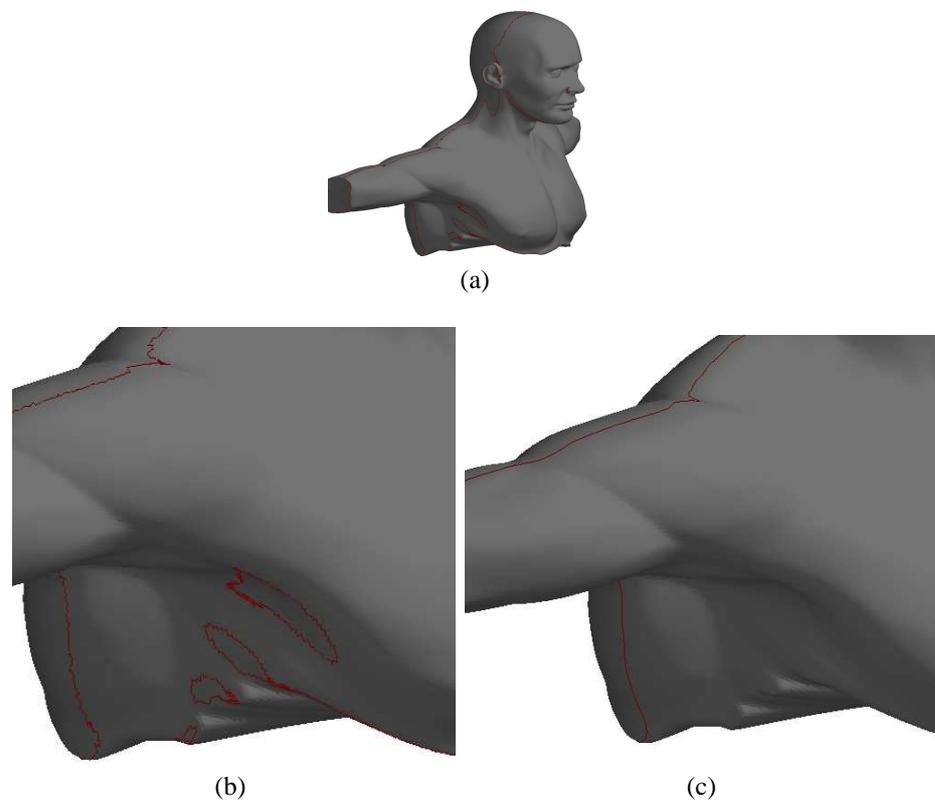


Figure 3.10: Silhouettes. (a) Silhouette computed on the model of a torso. (b) Zoomed-in view of the model with noisy silhouette (c) Simplified silhouette.

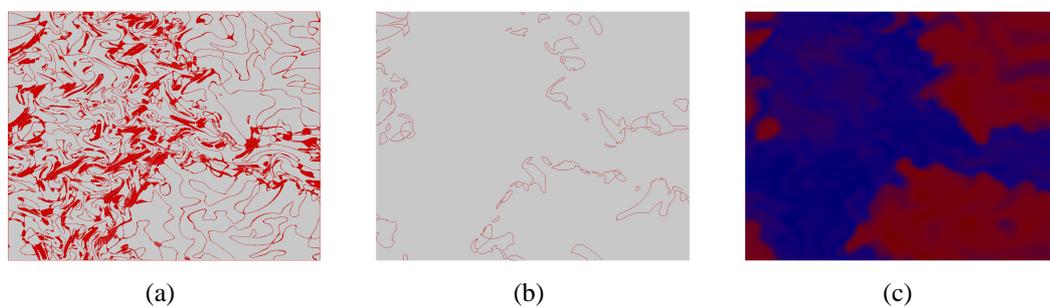


Figure 3.11: Combustion. (a) Jacobi set of H_2 and O_2 in the 64th time step. (b) Simplified Jacobi set. (c) Concentration of O_2 .

higher dimensional parts. During simplification, Jacobi set components within degenerate regions are automatically removed because they do not contribute to κ .

The dataset consists of the concentrations of H_2 (fuel) and O_2 (air) defined on a 600×600 grid for 67 time steps. We compute and simplify the Jacobi set for H_2 and O_2 at different time steps to identify the front of combustion. Combustion begins at regions where the fuel-air mixture is appropriate for ignition. The data is degenerate away from the front, thereby introducing noise in the Jacobi set.

Figure 3.11 shows the results for the 64th time step when the combustion is in its final stage. The simplified Jacobi set again appears at the front. Figure 3.11c shows the O_2 concentration. Blue signifies a low function value and red signifies a high function value. The front consists of the boundary of red regions, which is also traced by the simplified Jacobi set.

3.7 Conclusions

We have described an algorithm for simplifying the Jacobi set of two Morse functions. The algorithm is robust because it ensures minimal change to the relationship between the two functions (bounded change to the global comparison measure). The Jacobi set can be displayed at different levels of simplification due to the nature of our greedy algorithm. This allows for a representation of the Jacobi set with different levels of simplification (multi-level representation). Future work includes extending the algorithm to multiple functions and higher dimensions.

Chapter 4

Relation-aware Isosurface Extraction

In the previous chapter, we used the comparison measure to compute and simplify the Jacobi set. In this chapter, we describe a new technique that uses the comparison measure to identify important isosurfaces in multifield data.

4.1 Introduction

The design of interactive and useful techniques for multi-field data remains a challenging problem. Scientists hope to understand the underlying phenomena by studying the relationship between several quantities measured or computed over a domain of interest. Therefore, multi-field data is ubiquitous to all scientific studies.

Naturally, the design of analysis and visualization techniques for multi-field data will benefit by studying the relationship between fields as opposed to a focused study of inherent properties of individual fields. We follow this principle to develop a relation-aware method for exploring scalar multi-field data.

Identification of important isovalues of scalar fields is a well studied problem. Current approaches focus on individual scalar fields and study geometric properties of the isosurface like surface area or enclosed volume, or study the topological properties abstracted into a Reeb graph or contour tree. We study this problem in the context of multi-field data. Specifically, we

introduce a variation density function, whose profile is a good indicator of interesting isovalues of individual scalar fields in a multi-field dataset.

There is no unique definition for the relationship between functions in the literature. The alignment of gradients is widely used within the visualization community to quantify relationship between scalar fields [2, 3, 4]. We follow this approach and use the comparison measure proposed by Edelsbrunner et al. [3] to measure the relationship between different scalar fields.

For a given scalar field f and a sub-collection A of scalar fields from the multi-field data, the variation density function measures the relationship between scalar fields in A over isosurfaces of f . Similar to Edelsbrunner et al. [3], the variation density function quantifies the relationship between multiple scalar fields by comparing their gradients. Our hypothesis is that extrema and regions of rapid changes in the profile of the variation density function are indicative of interesting features or events in the data. Experiments on data from different applications indicate that our hypothesis is indeed true for these data sets.

Our contributions include a relation-aware approach to identification of interesting isovalues of a scalar field in a multi-field data set, a successful application of this approach to explore data from diverse application domains and a demonstration of the advantages over analyzing scalar fields in isolation. Central to the data exploration process is a variation density function that measures the relationship between scalar fields in the data. We derive links between the variation density and well understood measures like topological persistence and isosurface area statistics. We also describe a simple algorithm to compute an approximate profile of the variation density function, which provably converges to the true profile with increasing sample size. Finally, we show that our approach can be used for effective exploration of both simulation and measurement data from a wide variety of application domains.

4.1.1 Related Work

Bajaj et al. [8] introduced the popular *contour spectrum* as a method for exploring scalar fields by studying distributions of metric properties like area, volume, and their derivatives

and integrals. Early approaches to identification of interesting isovalues study the histogram of the scalar field [22, 23].

Carr et al. [7] showed that histograms suffered from many deficiencies because they were equivalent to nearest neighbor interpolant and suggest the use of isosurface statistics computed with higher quality interpolation. Scheidegger et al. [24] proposed an improved formulation of isosurface statistics by weighting it with the inverse gradient magnitude. This essentially means that the value of the statistic reaches infinity if the gradient vanishes. We fill this minor gap in the definition of the variation density function by excluding critical values. However, this does not affect the utility of the variation density function because we include the critical values while computing the variation density profile for piecewise linear input functions. Both Carr et al. and Scheidegger et al. mention an application of isosurface statistics to identification of interesting isovalues. Isosurface statistics considers geometric properties of an isosurface to determine its importance. For multi-field data, the importance of an isovalue additionally depends on the interaction between the different fields. In this regard, our method can be considered a generalization of their work to multi-field data. Section 4.5 describes this generalization in detail.

Structures like contour trees [25], and more generically Reeb graphs [26], provide an abstract representation of topological changes in isosurfaces of a scalar field as we sweep the domain in the direction of increasing / decreasing scalar value. The Reeb graph has been used as an interface for flexible extraction of individual components of interesting isosurfaces [27].

All the above methods are oblivious to other scalar fields in the data and hence do not consider relationships between fields. So, these methods may not be effective in the study of multi-field data.

Gosink et al. [2] present a method that allows visualization of interaction between three scalar fields by studying the correlation between two fields over isosurfaces of the third field. Their approach allows the classification of isosurfaces into two classes, primary and secondary,

but does not provide further information to allow the identification of a smaller set of interesting isovalues. They design the method to be used within the framework of Query-Driven Visualization, which benefits from user queries.

4.1.2 Outline

Section 4.2 reviews the necessary background on comparison measure, defines the variation density function, and describes its properties. Section 4.3 describes an algorithm to compute the variation density profile. Section 4.4 reports results of experiments on 2D, 3D, and time-varying multi-field data. Section 4.5 discusses some properties of the variation density function and an interesting variant. Section 4.6 concludes the paper.

4.2 Variation Density Function

The variation density function measures the relationship between multiple scalar fields over isosurfaces of one of the input scalar field. The relationship is quantified by the comparison measure introduced by Edelsbrunner et al. [3].

4.2.1 Comparison Measure

Let \mathbb{M} be a smooth compact n -dimensional Riemannian manifold. Let $F = \{f_1, f_2, \dots, f_k\}$ be a set of $k \leq n$ smooth real-valued functions defined on \mathbb{M} , $f_i : \mathbb{M} \rightarrow \mathbb{R}$. The *comparison measure*, for F , over a domain $D \subseteq \mathbb{M}$, is defined as the normalized integral

$$\kappa_D(F) = \frac{1}{\text{vol}(D)} \int_{x \in D} \|df_1 \wedge df_2 \wedge \dots \wedge df_k\|,$$

where $\text{vol}(D)$ is the volume of D and $df_1 \wedge df_2 \wedge \dots \wedge df_k$ is the wedge product of the k derivatives. $\kappa_{\mathbb{M}}(F)$ is called the *global comparison measure*. When D shrinks to a point $x \in \mathbb{M}$, we get the *local comparison measure*, $\kappa_x(F)$, in the limit. The product, $\kappa_D(F) \cdot \text{vol}(D)$,

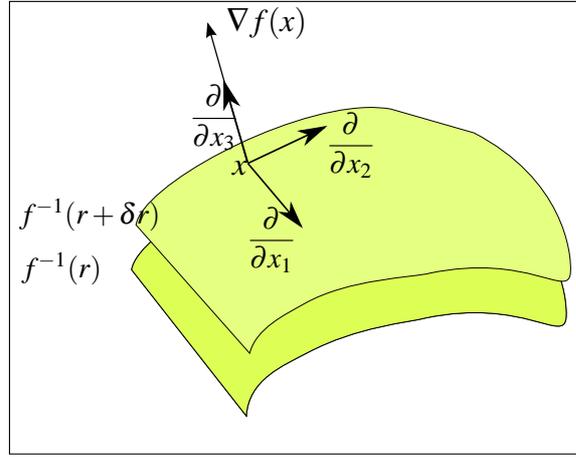


Figure 4.1: Isosurfaces shown at $f^{-1}(r)$ and $f^{-1}(r + \delta r)$ when $n = 3$. It is always possible to choose a local orthonormal coordinate system such that one tangent vector $\frac{\partial}{\partial x_3}$ is aligned with the gradient of f . From the definition of gradients, it follows that $dr = \|\nabla f(x)\| dx_3$.

can be used to quantify the relationship between the different functions in F . When $k = n = 2$ and \mathbb{M} is smoothly embedded in \mathbb{R}^3 with the standard Euclidean metric, $\kappa_x(F)$ is the length of the cross product of the two gradients at x :

$$\kappa_x(\{f_1, f_2\}) = \|\nabla f_1(x) \times \nabla f_2(x)\|.$$

4.2.2 Definition

For a smooth function $f : \mathbb{M} \rightarrow \mathbb{R}$, a real number c is a critical value if for some $x \in \mathbb{M}$, $\|\nabla f(x)\| = 0$ and $c = f(x)$. Let $\mathbb{P} \subseteq \mathbb{R}$ denote the set of non-critical, or *regular*, values of f .

Define a scalar function $\psi : F \times 2^F \times \mathbb{P} \rightarrow \mathbb{R}$ as

$$\psi(f, A, r) = \int_{x \in f^{-1}(r)} \frac{\kappa_x(A)}{\|\nabla f(x)\|} dS_x,$$

where dS_x is the $n - 1$ dimensional isosurface area element. We assume that all the functions in F have a finite number of critical values. Given a regular value r , we can therefore choose

an interval $I : [r, r + \delta r]$ that contains no critical values. We claim that

$$\begin{aligned} \int_I \psi(f, A, r) dr &= \int_{x \in f^{-1}(I)} \kappa_x(A) dV_x \\ &= \kappa_{f^{-1}(I)}(A) \cdot \text{vol}(f^{-1}(I)), \end{aligned} \quad (4.1)$$

where dV_x is the volume element. So, $\psi(f, A, r)$ can be considered as the density of the variation between functions in the set A . Hence, we call ψ the *variation density function*. We now prove (4.1) from first principles.

Consider a local coordinate system (x_1, x_2, \dots, x_n) at x such that the unit tangent vectors $(\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n})$ form an orthonormal basis. The volume element dV_x equals $dx = dx_1 dx_2 \dots dx_n$. Assume, without loss of generality, that the first $n - 1$ basis vectors lie on the tangent plane of $f^{-1}(r)$ at x and the last tangent vector is aligned with $\nabla f(x)$ (see Figure 4.1). We transform the coordinate system at x to $(x_1, x_2, \dots, x_{n-1}, f(x))$. The volume element in the new coordinate system is obtained by multiplying with the Jacobian determinant, which is equal to the length of the gradient $\|\nabla f(x)\|$. Therefore

$$dV_x = \frac{dx_1 dx_2 \dots dx_{n-1} dr}{\|\nabla f(x)\|}.$$

Now,

$$\begin{aligned} \int_I \psi(f, A, r) dr &= \int_I \int_{x \in f^{-1}(r)} \frac{\kappa_x(A)}{\|\nabla f(x)\|} dS_x dr \\ &= \int_I \int_{x \in f^{-1}(r)} \frac{\kappa_x(A)}{\|\nabla f(x)\|} dx_1 dx_2 \dots dx_{n-1} dr. \end{aligned}$$

Rewriting the double integral as a single integral over $f^{-1}(I)$ and using the above expression for dV_x , we get the desired equality in (4.1).

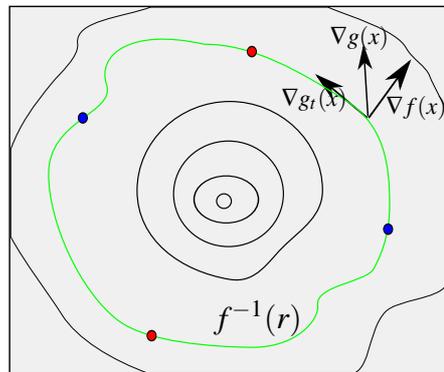


Figure 4.2: Isocontours of f when $n = 2$ and $F = \{f, g\}$. The isocontour $f^{-1}(r)$ shown in green, has critical points of the restricted function g_* , at the points shown in red (maxima) and blue (minima). The ψ function is the sum of persistence values of the critical points. It also captures the total variation of g_* on the isocontour.

4.2.3 Variation Density and Total Variation

In this section, we motivate the use of variation density by showing that it captures the total “variation” of a function restricted to an isosurface of the other. Specifically, we show that, for the special case of smooth functions f and g defined on a 2-manifold, the variation density is equal to the difference between the sum of values at maxima and minima of g restricted to an isocontour $f^{-1}(r)$. However, it is not clear how to extend this result to the case of multiple functions defined on a higher dimensional manifold.

Let r be a regular value of f . The isocontour $f^{-1}(r)$ is a smooth curve embedded in \mathbb{M} . So, we have

$$\begin{aligned} \psi(f, \{f, g\}, r) &= \int_{x \in f^{-1}(r)} \frac{\|\nabla f(x) \times \nabla g(x)\|}{\|\nabla f(x)\|} dl_x \\ &= \int_{x \in f^{-1}(r)} \|\nabla g_t(x)\| dl_x, \end{aligned} \quad (4.2)$$

where $\nabla g_t(x)$ is the component of $\nabla g(x)$ along the tangent to $f^{-1}(r)$ at x , and dl_x is the length element of $f^{-1}(r)$ at x (see Figure 4.2). Let g_* be the function obtained by restricting the domain of g to $f^{-1}(r)$. The derivative of g_* vanishes at a critical point. Critical points are

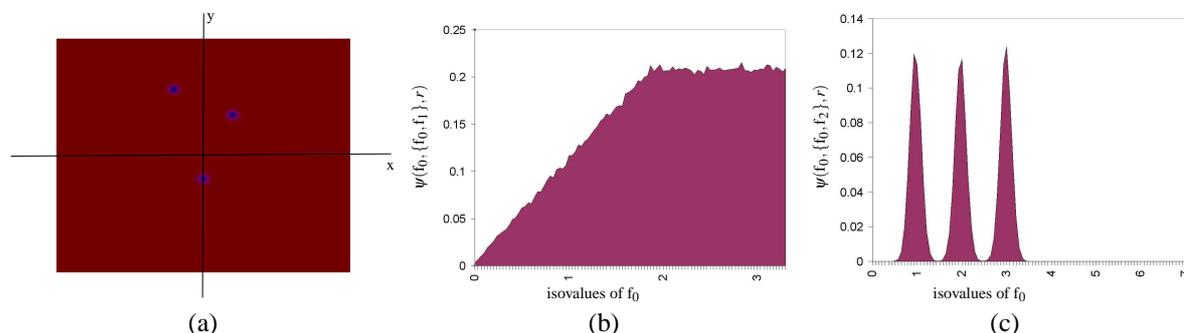


Figure 4.3: (a) Color map of function f_2 . Blue and red regions indicate low and high function values respectively. (b) The variation density profile of f_0 , $\psi(f_0, \{f_0, f_1\}, r)$ with $f_1(x) = \|x - 2\|$. The jagged boundary of the shaded region is an artifact of using a discrete domain for computation. (c) The variation density profile of f_0 , $\psi(f_0, \{f_0, f_2\}, r)$, showing peaks at depressions of f_2 .

either maxima or minima assuming the second derivative of g_* does not vanish at such points. Applying the fundamental theorem of calculus to each region of $f^{-1}(r)$ where g_* is monotone, we rewrite the integral of $\|\nabla g_t(x)\|$ over the isocontour $f^{-1}(r)$ as the difference between the sum of function values at maxima and minima of g_* . In other words, if C is the set of critical points of g_* , then

$$\int_{x \in f^{-1}(r)} \|\nabla g_t(x)\| dl_x = 2 \sum_{v \in C} \text{sign}(v) g_*(v), \quad (4.3)$$

where $\text{sign}(v)$ is either $+1$ or -1 depending on whether v is a maximum or minimum, respectively. Thus, ψ is equal to the total variation of g_* over the isocontour $f^{-1}(r)$.

The *sub-level set* of a real value s is the union of pre-images of all real values less than or equal to s . Consider the sub-level sets of g_* as we sweep $f^{-1}(r)$ in the direction of increasing value of g_* . New components are created at local minima of g_* . Components of the sub-level set merge at all maxima except for the global maximum where the sub-level set is equal to $f^{-1}(r)$. We represent each sub-level set component by its oldest minimum. When a merge happens at a maximum, we pair the maximum with the younger of the two minima representing the two merging components. The global maximum is paired with the global minimum. The

persistence of a critical point is equal to the absolute difference in function values between the critical point and its pair. Persistence of a critical point pair represents the lifetime of a feature, *i.e.*, the time between the creation and destruction of a component in $f^{-1}(r)$ during the sweep process. Long-living components that have higher persistence values are considered to be more important.

In recent years, the notion of persistence has been used to measure, order, and simplify features [14, 17, 28, 29, 30]. This measure has been shown to be stable in the presence of noise assuming a bottleneck metric and the functions are tame [31]. From Equation (4.2) and Equation (4.3), we know that

$$\psi(f, \{f, g\}, r) = 2 \sum_{v \in C} \text{sign}(v) g_*(v). \quad (4.4)$$

Since every critical point is counted twice in the above expression, $\psi(f, \{f, g\}, r)$ is equal to the sum of persistence values of all critical points of g_* . The variation density function ψ , therefore, represents the total importance of all sub-level set components of g_* in $f^{-1}(r)$. This equality also suggests that we can expect the variation density function to be insensitive to small perturbations in g . This is because a small perturbation applied to the function implies a small perturbation applied to the restriction of g_* . Now, g_* and its perturbed versions are close to each other under the L_∞ metric, which implies that the persistence values of their critical points, and hence their sum, are close to each other. Note that the function is tame because we assume that it has a finite number of critical values.

Finally, note that the integral of the expression in Equation (4.4) over all isovalues of f is equal to the global comparison measure as shown previously by Edelsbrunner et al [3].

4.2.4 Variation Density Profile

We are interested in the plot of variation density for a given scalar field f and a subset A of scalar fields. The observation that the integral of the variation density over all isovalues

is the global comparison measure motivates us to study the plot of variation density against isovalues. We can consider the profile of the variation density as a plot of the contribution to the global comparison measure from the level sets of the scalar field. Given k scalar fields, only a few of the $k2^k$ possible plots are interesting. This choice of A and f is typically determined by the application. Prior knowledge of potential interaction between the scalar fields can help us make an informed choice. Each plot can provide cues that help in identifying interesting isovalues. The following examples are aimed at providing intuition behind the use of the variation density profile.

Consider the following analytic functions defined on \mathbb{R}^2 .

$$\begin{aligned} f_0(x) &= \|x\|, \\ f_1(x) &= \|x - a\|, \\ f_2(x) &= -(G_{a_1}(x) + G_{a_2}(x) + G_{a_3}(x)). \end{aligned}$$

Where $x, a, a_1, a_2, a_3 \in \mathbb{R}^2$ and $G_{a_i}(x)$ is a Gaussian with a low standard deviation centered at a_i . The isocontours of f_0 and f_1 are circles centered at origin and the point a respectively. Consider an isovalue $r < \|a\|$ of f_0 . The value of $\psi(f_0, \{f_0, f_1\}, r)$ can be calculated from using Equation (4.4) to be $4r$. If $r \geq \|a\|$, we have $\psi(f_0, \{f_0, f_1\}, r) = 4\|a\|$. The function ψ , therefore, increases linearly with r till the isovalue $\|a\|$ and then becomes constant.

If we consider f_0 to be elevation and f_1 to be atmospheric pressure, ψ would tell us that the variation in pressure at all points with the same elevation increases linearly till height $\|a\|$ and remains constant for higher elevations. The pressure depression at elevation $\|a\|$ is captured by a knee in the graph of ψ , see Figure 4.3b. Note that the corrected isocontour perimeter statistic [24] would assign the same value in the statistic graph for each isovalue of f_0 . This follows from the fact that the ratio of the perimeter of an isocontour to the length of the gradient of f_0 on the isocontour is the same for all isocontours.

The function f_2 has three depressions at distances $\|a_1\|$, $\|a_2\|$ and $\|a_3\|$ from the origin.

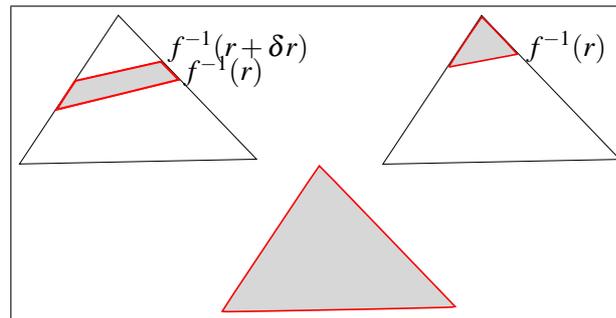


Figure 4.4: Computation of ψ for a two dimensional simplicial complex. The shaded region indicates the area corresponding to a bin $[r, r + \delta r]$. This region is a small strip if the bin is contained within the range of f restricted to the triangle. The entire triangle is shaded if the bin contains the range of f restricted to the triangle or if the triangle is degenerate. If the bin and the range of f have a non-empty intersection but do not contain each other, then the shaded region contains one or two vertices of the triangle.

Figure 4.3a shows a color map of f_2 with depressions (blue regions) centered at distances 1, 2 and 3 from the origin. The function f_2 is nearly constant at all points far away from the depressions. The variation of f_2 on an isocontour of f_0 is nearly zero if the isocontour does not pass through any of the depressions. The variation is maximum on isocontours of f_0 passing through the depressions. This results in peaks in the variation density profile (see Figure 4.3b). In both the examples, select isovalues of f_0 are found to be interesting only after studying the relationship between functions.

4.3 Computation

In this section, we describe the computation of ψ when F is a set of piecewise-linear functions and \mathbb{M} is represented by an n -dimensional simplicial complex (see Chapter 2).

The gradient of a piecewise-linear function f_p is well defined in the interior of a simplex and is a constant vector because f_p is linear within the simplex. The gradient vanishes inside a simplex iff the function values at all vertices of the simplex are equal. Note that a constant gradient implies that the local comparison measure is also constant in the interior of a simplex.

For a smooth function f , we first divide the range of f into a fixed number of intervals

called bins. For an interval $I = [r, r + \delta r]$, define

$$\bar{\psi}(f, A, I) = \frac{\int_{f^{-1}(I)} \kappa_x(A) dx}{|I|}, \quad (4.5)$$

where $|I|$ is the length of the interval I . The function $\bar{\psi}(f, A, I)$ is well defined even if I contains a critical value. Note that

$$\begin{aligned} \psi(f, A, r) &= \lim_{\delta r \rightarrow 0} \frac{\int_r^{r+\delta r} \psi(f, A, r) dr}{\delta r} \\ &= \lim_{\delta r \rightarrow 0} \frac{\int_{f^{-1}(I)} \kappa_x(A) dV_x}{\delta r} \end{aligned}$$

Therefore, in the limit, when $|I| \rightarrow 0$, $\bar{\psi}$ converges to $\psi(f, A, r)$ at a regular value r .

For a piecewise-linear function f_p , we compute the integral in Equation (4.5) as a summation:

$$\bar{\psi}(f_p, A, I) = \frac{1}{|I|} \sum_{\sigma \in K} \kappa_\sigma * \text{vol}(\text{interior}(\sigma) \cap f_p^{-1}(I)),$$

where κ_σ is the value of κ_x for any $x \in \text{interior}(\sigma)$ (see Figure 4.4). Note that for piecewise-linear functions, the local comparison measure is a piecewise-constant function. Therefore, κ_x is the same for any $x \in \text{interior}(\sigma)$. The procedure COMPUTEPSI computes the variation density profile for a given bin width h . The procedure is easily parallelizable because the computation for each simplex is independent of other simplices. The time required for a simplex inside the outer loop depends on the range of the function restricted to it and h . The worst case complexity is therefore $O(mn)$, where m and n are the number of bins and simplices respectively.

Procedure COMPUTE $\Psi_I(f_p, A, h)$

 Initialize $\bar{\psi}(f_p, A, I) \leftarrow 0$ for all bins I
for each simplex $\sigma \in K$ **do**
 $R \leftarrow$ range of f_p restricted to σ
for each bin I such that $R \cap I \neq \emptyset$ **do**

$$\bar{\psi}(f_p, A, I) \leftarrow \bar{\psi}(f_p, A, I) + \frac{\kappa_\sigma * \text{vol}(f_p^{-1}(R \cap I) \cap \text{interior}(\sigma))}{h}$$

end for
end for

4.4 Applications

We study a variety of data using the variation density function. Our implementation works directly on simplicial complexes. If the input domain is available as a rectilinear grid, we first subdivide it into simplices by inserting diagonals and analyze the corresponding piecewise linear function. Area and volume are computed using the QHull library (<http://www.qhull.org>). We have also parallelized the computation using OpenCL ¹ (<http://www.khronos.org/ocl/>). The software is available for free from <http://vgl.serc.iisc.ernet.in/software/software.php?pid=002>.

We use a fixed number of bins (100 or 200) in all our experiments. We focus on local maxima, minima and regions of steep gradients in the profile in order to identify potentially interesting isosurfaces. In all experiments, we compare our result with the isosurfaces identified using the corrected isosurface area statistic [24].

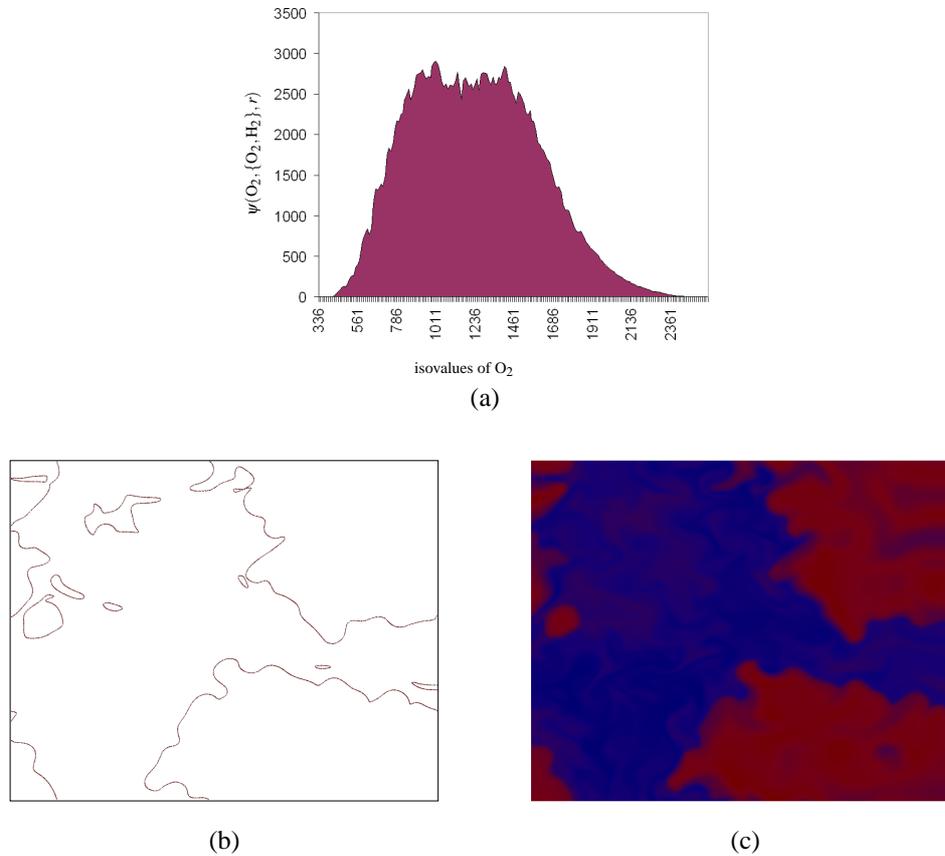


Figure 4.5: Profiling isocontours of oxygen during a combustion simulation. (a) ψ computed with $f = \text{O}_2$ and $A = \{\text{O}_2, \text{H}_2\}$. (b) Isocontour of O_2 at 1010, the global maximum of ψ . Isocontours of oxygen in the range (800, 1400) belong to the front of combustion. (c) Color map of oxygen with red and blue areas indicating high and low concentrations respectively.

4.4.1 2D Combustion

We first experiment on a 2D combustion simulation data². Hydrogen fuel at 300K is mixed with an oxidizer (21% oxygen) at 1200K. The goal of the simulation is to study the influence of turbulence on the different phases of combustion. When compressed, the fuel ignites at multiple spots because of the inhomogeneity in the air-fuel ratio. Depending on the air-fuel ratio, the flame either propagates in an outward direction from the ignition spot or burns out [19, 32].

¹Shantanu Chaudhary wrote the implementation using OpenCL for computing the variation density profile on GPUs.

²We would like to thank Valerio Pascucci and Jackie Chen for providing the combustion data.

The combustion is simulated on a plane over 67 time steps. The input data comprises of three scalar fields defined on a 600 x 600 grid for 67 time steps. The value of the first field at each point indicates the progress of combustion at the point. The concentrations of oxygen (O_2) and hydrogen (H_2) are the other fields.

The concentrations of H_2 and O_2 are nearly constant away from the front of combustion. Also, the gradients of the two functions are aligned in these regions. The comparison measure is therefore zero in these regions. The variation density of oxygen or hydrogen will therefore have non-zero values only for isocontours passing through the front.

For this experiment, we consider the 64th time step as our input domain. The combustion is in its later stages in this time step. We profile the isocontours of oxygen considering its relationship with the fuel (hydrogen). This is accomplished by choosing $A = \{O_2, H_2\}$. The variation density profile is shown in Figure 4.5a.

We observe from the profile that it increases to a maximum when the oxygen level is approximately 800 and remains high till the oxygen level is approximately 1400. We notice a gradual decline for higher isovalues. The isocontours of oxygen in this range (800, 1400) belong to the front of the combustion. However, this information cannot be directly inferred from the isosurface statistic. Figure 4.5b shows the isocontour of O_2 at the value 1010, the global maximum of ψ . The front is the region where the fuel is actively burning.

The scientists who designed the simulation commented that the isocontour based segmentation of the ignition region or a burned out/extinction hole is useful in studying and understanding the nonlinear coupling that governs ignition and extinction. The shape and size of the segmented region and the correlation between the multitude of scalar fields computed within the segment play an important role in the study. We also observed that the level of detail of the front was higher at O_2 concentration ~ 800 compared to ~ 1400 .

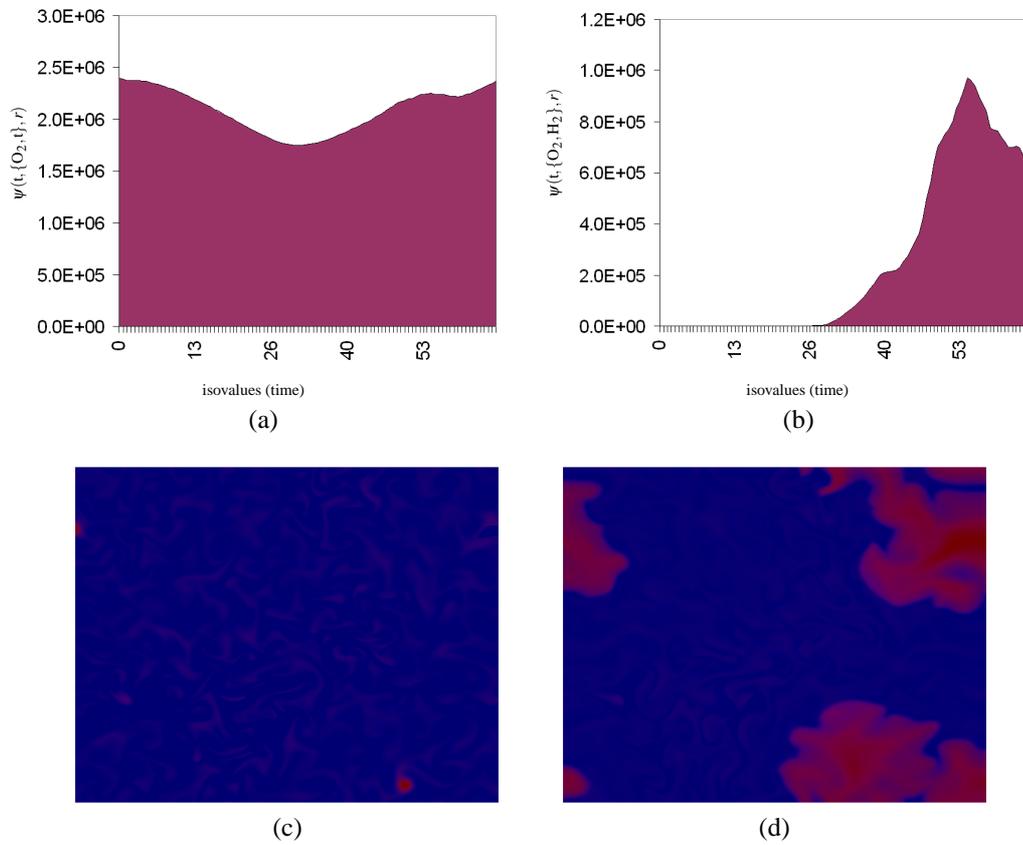


Figure 4.6: Time varying combustion. (a) Variation density profile with $f = t, A = \{t, O_2\}$. (b) Profile of ψ with $f = t, A = \{H_2, O_2\}$. (c) Ignition: color-mapped image of the scalar field `prog`, which measures the completion of combustion. Red regions indicate high values and blue regions indicate low values. The distribution of `prog` in the 28th time step indicates the regions where the fuel is ignited. (d) Burning: The distribution of `prog` in the 52nd time step.

4.4.2 Time Varying Combustion

Next, we show the application of the variation density function to time-varying data. We consider the time varying combustion data described in the previous experiment as a three-dimensional data with time t defined as an additional scalar field. The fuel consumption rate at a point in a time step can be used to measure the progress of combustion at the point. This information is available as a scalar function `prog`.

The goal of this experiment is to identify important time phases of the combustion process. The relationship that O_2 has with time changes during the important phases of combustion. For

example, on every time slice before ignition, the concentration of oxygen is nearly constant everywhere and on ignition, oxygen begins to be consumed at regions of ignition. We therefore profile the isosurfaces of time with $A = \{t, O_2\}$ (see Figure 4.6a). The profile successfully captures the ignition and the burning phases of the combustion process. The time steps $t = 27$ to $t = 35$ in the data correspond to the ignition process. This is captured as a trough in the profile. The burning phase ($t=50$ to $t=55$) is also captured by a maximum in the plot. Ignition and burning are indeed considered to be the two important phases of this combustion process [32].

The interaction between O_2 and H_2 can also be considered to determine the different phases. There is no real interaction between O_2 and H_2 before ignition. We plot ψ with $A = \{O_2, H_2\}$ hoping to find more information (see Figure 4.6b). The information extracted from this profile is essentially the same compared to the profile ψ with $A = \{t, O_2\}$. The profile begins to increase from zero during the ignition phase and reaches a global maximum during the burning phase. Isosurface statistics considers only geometric properties of a time slice and hence would not be able to detect any of the above phases. For example, the corrected area statistic would give equal importance to each isovalue and hence the plot would be a horizontal line.

The developers of the combustion simulation noted that it is desirable to identify and track transient and intermittent events like auto-ignition and extinction³. They comment that our approach of studying the relationship between the air-fuel mixture over the non-local geometry of the flame front is a new idea and could help attain further insights into flame interactions.

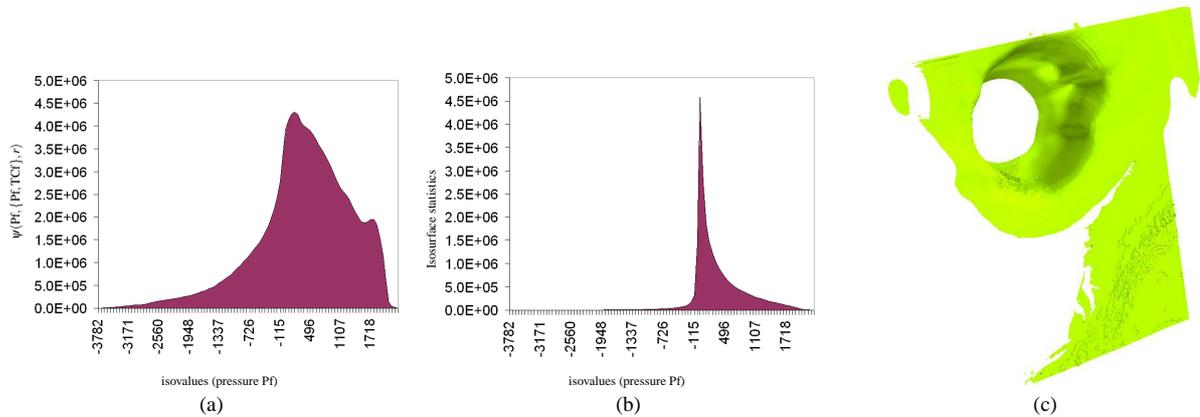


Figure 4.7: Hurricane Isabel. In regions of low correlation between the fields (-100 pascals, 20 pascals) (a) the variation density function and (b) isosurface statistics behave similarly. The two values are nearly equal at isovalue 7 pascals. (c) Isobar of the hurricane at 7 pascals.

4.4.3 Hurricane Isabel

Hurricane Isabel was a strong hurricane that struck the west Atlantic region in September 2003. We consider a simulation of this event [33]⁴. The domain is a 3D rectilinear grid of size $500 \times 500 \times 100$ corresponding to a physical scale of $2139km \times 2004km \times 19.8km$. Eight scalar fields are defined over this domain. This data is defined for 48 time steps corresponding to an actual time of 48 hours. For experimental purposes, we look at only pressure (Pf) and temperature (TCf).

We study the isosurfaces of pressure at the first time step with $A = \{Pf, TCf\}$. During the initial phase of the hurricane, the eye of the storm was located in the ocean. The swirling motion around the eye corresponded to a low pressure region (-100 pascals, 20 pascals). Temperature and pressure have low correlation in this region [4]. The variation density profile shows an exponential increase for the isobars corresponding to low pressure (Figure 4.7a).

A natural question to ask is “Under what conditions do the isosurface statistics and the variation density function produce similar results?”. In areas of low correlation, we observe

³We thank Jackie Chen and Ajith Mascarenhas for their help with interpreting the results of our experiments on the combustion data.

⁴Hurricane Isabel data was produced by the Weather Research and Forecast (WRF) model, courtesy of NCAR and the U.S. National Science Foundation (NSF)

that the isosurface statistics plot (see Figure 4.7b) behaves similarly. Both exhibit an exponential increase in the region of low pressure. In fact, at approximately 7 pascals pressure, we observe that the two plots have nearly equal values. These experimental observations lead us to believe that when the input fields have a low correlation, the variation density function has no added advantage over isosurface statistics.

4.4.4 Universe Simulation

In the fourth experiment, we consider the simulation of ionization front instability in the universe [34]⁵. The input domain is a $600 \times 248 \times 248$ rectilinear grid which is equivalent to a physical volume of $0.6 \text{ parsec} \times 0.25 \text{ parsec} \times 0.25 \text{ parsec}$. The simulation is done over 200 time steps corresponding to 25.37 thousand years. The data has ten different simulated scalar fields: particle density, temperature (TCf), and eight chemical species including gaseous hydrogen (H_2), ionized hydrogen (H^+) and ionized helium (He^+).

The ultraviolet radiations from stars ionize hydrogen and oxygen present in space. This ionization process slows down the photons, which now proceed at a much slower pace behind a radiation wall known as the ionization front. This front separates the hot gases ($> 20000\text{K}$), which are in an ionized state, from the ambient space at 72 K.

We study the impact of each of the chemical species on the importance of isotherms. We first study the effect of H^+ . Since hydrogen is in the ionized state, we expect to find the relevant isotherms at high temperatures. This is indeed the case (see Figure 4.8b). The profile peaks in the temperature range 14000-16000K, which is the temperature range in which hydrogen is ionized. We get similar results for ionized helium ($A = \{\text{TCf}, \text{He}^+\}$) (Figure 4.8c). The temperatures relevant for gaseous hydrogen (H_2) (Figure 4.8d) were found to be 2000-15000K, after which the plot goes to zero. This is in accord with the known fact that hydrogen is typically in the ambient state ($>72\text{K}$) or shocked state ($>2000\text{K}$). Above 15000K, hydrogen is

⁵Universe simulation data was produced by Daniel Whalen at Los Alamos National Labs and Michael L. Norman at San Diego Supercomputer Center.

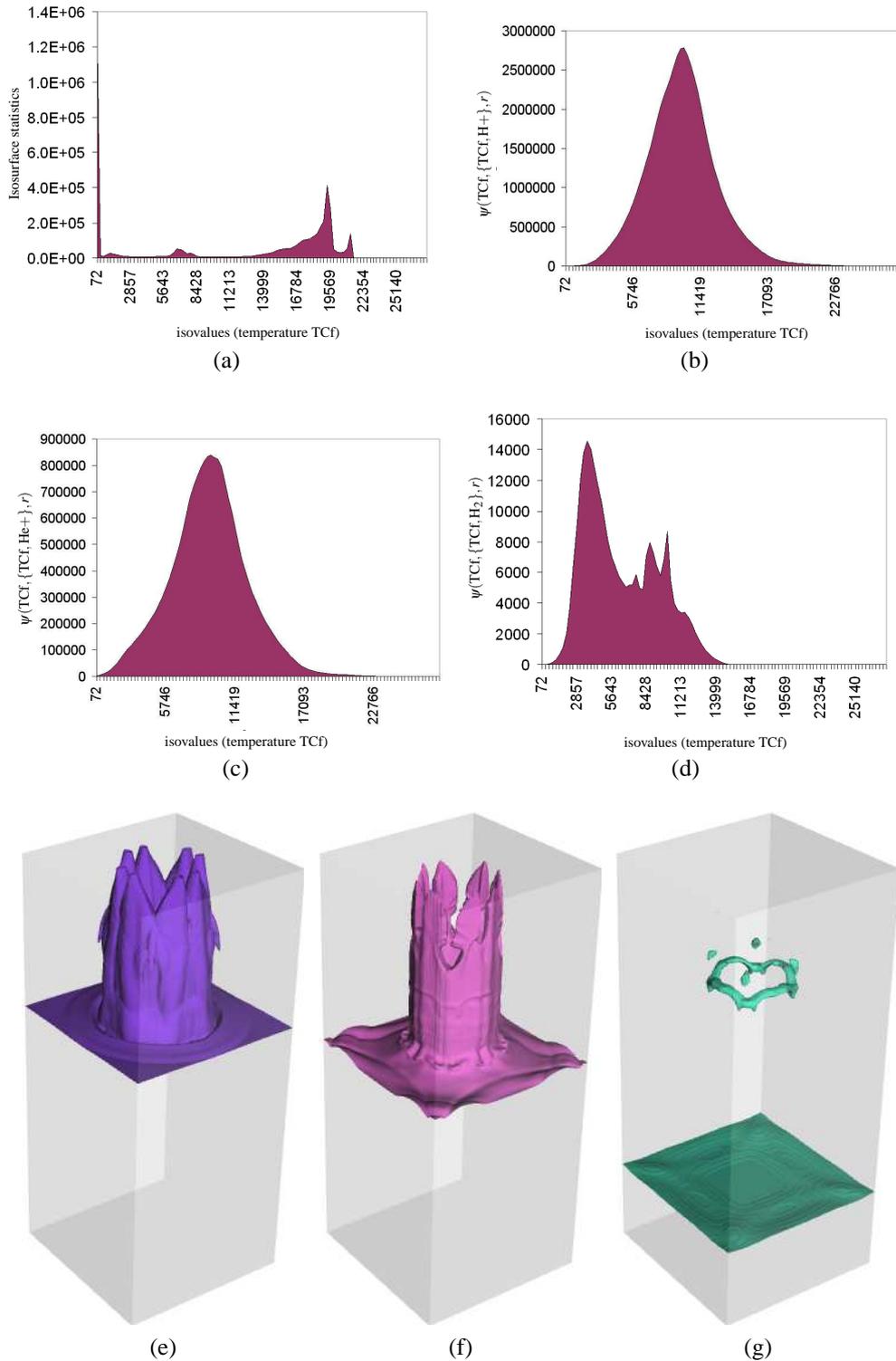


Figure 4.8: Universe Simulation. (a) Isotherm statistics. (b) The profile of ψ plotted with $A = \{\text{TCf}, \text{H}^+\}$. Most of the area under the curve is centered around the ionization temperature of hydrogen. Similar results are seen in (c) when $A = \{\text{TCf}, \text{He}^+\}$. (d) The profile of ψ with $A = \{\text{TCf}, \text{H}_2\}$. Hydrogen in the temperature range 72-14000K is in either cool or shocked states. (e) Isotherm of universe at 3000K which lies in the range where hydrogen is in shocked state. (f) Isotherm of universe at 14500K which lies in the range where hydrogen and helium are ionized. (g) Isotherm of universe at 19500K which lies in the range where hydrogen is already ionized and there is no other significant event.

primarily in the ionized state. Isotherms in Figures 4.8(e-f) correspond to isovalues identified from the variation density profile. These isovalues belong to ranges where hydrogen is in a shocked state (Figure 4.8e), hydrogen and helium are ionized (Figure 4.8f), and where there is no specific interaction between temperature and the different elements (Figure 4.8g). The geometry of the isotherm has no particular interpretation to the best of our knowledge.

4.5 Discussion

The derivation of isosurface statistics by Scheidegger et al. [24] may also be extended to develop a relation-aware statistic. In the case of three dimensional domains, isosurface statistics considers the volume enclosed by the isosurfaces $f^{-1}(r)$ and $f^{-1}(r + \delta r)$ as δr approaches zero. Normalizing this volume by the volume of the manifold, we get a probability density function that measures the probability that the scalar field assumes values between r and $r + \delta r$ as δr approaches zero. Clearly, the profile of this probability density function is the same as the isosurface statistic. However, the notion of a probability density function can be extended to two fields.

When two fields are available, we may consider the joint probability density (JPD). Rajwade et al. [35] use the JPD for two scalar fields in the context of computing mutual information and solving the image registration problem. The scalar fields are essentially greyscale of the two images that are to be registered. They show that the JPD equals

$$p(\alpha_1, \alpha_2) = \int_{\{x|f(x)=\alpha_1\} \cap \{x|g(x)=\alpha_2\}} \frac{dx}{\|\nabla f(x) \times \nabla g(x)\|},$$

where f and g are the scalar fields, and α_1 and α_2 are isovalues of f and g respectively. The JPD is essentially the continuous scatterplot recently introduced by Bachthaler and Weiskopf [36]. Note that this integrand is equal to the inverse of the local comparison measure κ_x , which suggests a direct extension to multiple fields.

We also observe that the isosurface area statistic [7] and the corrected statistic [24] can be

derived as special cases of the variation density function. If the set A contains a single element f , the scalar field under consideration, then the local comparison measure $\kappa_x(\{f\}) = \|\nabla f(x)\|$. This implies that the variation density function

$$\psi(f, \{f\}, r) = \int_{x \in f^{-1}(r)} dS_x,$$

which is exactly the isosurface area statistic derived by Carr et al. Now, consider the case when κ_x is a constant function, which essentially means that we have no additional information on the relationship between the scalar fields. In this case, the variation density function reduces to the corrected isosurface statistic.

The derivation in Section 2.3 indicates that the variation density function is not likely to be susceptible to noise, especially when the dimension of the domain is less than three. The derivation, however, extensively utilizes the property that regions in an isocontour can be broken into monotone paths of the restricted function g_* resulting in a closed form expression for the integrals. It is unclear if such an approach can be extended to higher dimensional domains.

4.6 Conclusions and Future work

We have introduced a variation density function ψ to profile isosurfaces based on relationships between different scalar fields in multi-field data. We also described an algorithm to compute the profile. The fact that ψ captures significant information that is typically not captured by isosurface statistics is evident from our experiments with several data sets from diverse real-world applications. We also conjecture that for fields with low correlation, ψ may be no better than isosurface statistics.

We list the following problems as future work:

- Characterizing the link between persistence and variation density in higher dimensions.

- Extension of our results to arbitrary number of scalar fields. Currently, the number of fields that can be compared (*i.e.* the size of the set A) is bounded by the dimension of the domain. This is primarily because the comparison measure considers the alignment of gradients of the fields to determine relationships. If the number of fields is greater than the dimension of the domain, the gradients become linearly dependent and hence the comparison measure is zero everywhere. One solution is to consider only a subset of fields at a time and then collect the different statistics together in a well defined way. In Chapter 5, we describe a new multifield comparison measure that allows the number of fields to be compared to exceed the dimension of the domain.
- Extending the definition of the variation density function to vector fields will be a challenging task because the comparison measure cannot trivially be extended to compare vector fields. We define a new multifield comparison measure that can be extended to multiple vector fields in Chapter 5.
- It would be interesting to see if single scalar fields can be studied more effectively using our approach. This would involve identifying suitable derived fields that can be used to profile the input scalar field.

Chapter 5

Multifield Comparison Measure

In Chapters 3 and 4, we noticed that the local/global comparison measure could compare scalar fields only if the number of fields did not exceed the dimension of the domain. In this chapter, we define a new gradient-based multifield comparison measure that can compare an arbitrary number of scalar fields.

5.1 Introduction

Data from present day simulations and observations of physical processes often consists of multiple scalar and vector fields. Studying the interactions between the fields is pivotal to understanding the underlying phenomenon.

Single scalar fields are typically studied using techniques like isosurfacing, direct volume rendering and contour trees [22, 25, 27, 37, 38]. When visualizing multiple scalar fields, the above methods can be used separately on each field and visualized side by side or as overlays. The relationships and interactions that exist between the fields are often not captured by such methods. Simultaneous visualization of all the fields facilitates the understanding of interactions and relationships between them. This can be accomplished by employing a comparative approach to capture the relationships between variables.

5.1.1 Proposed Approach

We present a new gradient-based comparison measure for scalar fields that is applicable on an arbitrary number of scalar fields defined on a manifold. The measure captures the extent of alignment of the gradient vectors at a point. The distribution of the measure over the domain provides key insights into the interaction between input fields. The measure satisfies various desirable mathematical properties, can be computed efficiently, and is practically useful for studying relationships between multiple scalar fields. We apply this measure for analyzing a hurricane simulation data set and a global climate simulation data set. The analysis helps explain various known meteorological and climatic phenomena. We also demonstrate the effective use of an aggregated version of the measure to the study of a combustion simulation data set.

The main contributions of this chapter are :

- A new multifield comparison measure to capture interactions between multiple scalar fields defined on an n -dimensional domain,
- Theoretical results that establish the robustness of the measure by showing its insensitivity to noise in the scalar fields,
- An algorithm to compute the measure efficiently, and
- Real world applications to demonstrate the effectiveness of the measure in studying interactions between scalar fields in physical phenomena and an extension to vector fields.

5.1.2 Related Work

A popular approach to visualizing multiple fields is to combine them into a single value and then render the combined volume [39,40]. Woodring et al. [41] propose that the data fields should be rendered together within the same space for user comparison. They use set operators to combine the different fields into a single field that extracts the interesting portions of the

data. These set operators can either combine the color values of the input fields or directly apply the operation in data space. Though combining volumes shows important parts of the data, the interactions between the different variables that are of importance to the domain scientists are not captured. For multifield time varying data, Lee et al. [42] propose a linear time algorithm to extract trend relationships among variables based on studying the change of variables over time and how these changes are related among different variables. Features in multifield data have been extracted using techniques like scatter plots [36] and variation density plots (see Chapter 4).

Multifield data have also been studied using statistical methods. One important work in this area uses the local statistical complexity [43] to identify features which may exhibit the same behavior in the future. Features are identified as complex if the probability that they occur again is low. In a later work, Jänicke et al. [44] improve the accuracy and efficiency of computing the local statistical complexity.

The relationship between the different scalar fields is popularly captured with the help of correlation measures. Sauber et al. [4] use two different techniques to compare different scalar fields at a point. One of them uses the alignment of gradients of the fields and also their magnitudes as a criterion to measure similarity. When the number of fields exceed two, pairwise similarity is computed and the least value is considered. This would detect regions where two of the fields are highly correlated. An obvious limitation of this approach is that two fields with low correlation would result in the other fields of the data to be ignored. In the same paper, the authors also describe a local correlation coefficient to detect linear dependencies between the scalar fields. The advantage of this method is its insensitivity to scaling of the data fields. It also has the same limitation as the first approach. Gosink et al. [2] also use correlation fields to study the interactions between the different variables in multi-field data. The inner product of the gradients of two fields of interest is computed over principle level sets of a third field. They use this approach to study combustion in methane and hydrogen. A limitation with using the inner product of the gradients is that only two fields can be compared.

Edelsbrunner et al. [3] also employ a gradient-based approach to measure relationships between scalar fields. In their work, they introduce a measure to compare multiple scalar fields both locally at a point as well as over a region of the domain. In the case of three dimensional Euclidean space and two fields, they show that the measure at a point reduces to the length of cross product of the gradients of the fields. This measure, though useful, has a limitation that the number of scalar fields that can be compared cannot exceed the dimension of the domain.

In this chapter, we also explore a gradient-based approach to compare scalar fields locally at a point. However, our method is not limited by the number of fields that can be compared unlike previous approaches. Our method also extends to time-varying scalar fields and to vector fields. Further, the measure is provably robust to noise in the input fields.

5.1.3 Outline

The rest of the chapter is organized as follows. In Section 5.2, we define the multifield comparison measure and prove its robustness and other properties. We motivate the use of the measure and explain its working in Section 5.3. Computation of the measure is described in Section 5.4. We describe several applications of the measure in Section 5.5. In Section 5.6, we discuss the limitations of the multifield comparison measure and its insensitivity to noise in a real world data. We conclude the chapter in Section 5.7.

5.2 Multifield Comparison Measure

In this section, we introduce a gradient-based comparison measure for multiple scalar functions. The measure is defined as the norm of a matrix comprising the gradient vectors of the different functions. We first define the matrix norm before defining the measure and listing and proving its properties.

5.2.1 Matrix Norm

Let A be a $m \times n$ matrix of real numbers. The *norm* of the matrix A , denoted as $\|A\|$, is defined as

$$\|A\| = \max_{\|x\|=1, x \in \mathbb{R}^n} \|Ax\|,$$

where $\|x\|$ represents the Euclidean norm of vector x [45]. We list four properties of the matrix norm that we will use later to prove key properties of the comparison measure. In particular, if A and B are matrices of real numbers, then

1. $\|A\| > 0$ if $A \neq 0$ and $\|A\| = 0$ iff $A = 0$.
2. For $\alpha \in \mathbb{R}$, $\|\alpha A\| = |\alpha| \|A\|$.
3. $\|A + B\| \leq \|A\| + \|B\|$ and $\|A - B\| \geq \left| \|A\| - \|B\| \right|$
4. $\|AB\| \leq \|A\| \|B\|$.

5.2.2 Comparison Measure

Let \mathbb{M} be a compact Riemannian manifold of dimension n . Let (x_1, x_2, \dots, x_n) be a local coordinate system such that the unit tangent vectors form an orthonormal basis with respect to the Riemannian metric. Let $F = \{f_1, f_2, f_3, \dots, f_m\}$ be a set of smooth functions defined on the manifold. The derivative at a point $p \in \mathbb{M}$ is written as a matrix of partial derivatives,

$$dF(p) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(p) & \cdots & \frac{\partial f_1}{\partial x_n}(p) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(p) & \cdots & \frac{\partial f_m}{\partial x_n}(p) \end{bmatrix}$$

We define the *multifield comparison measure* η_p^F at point p as the norm of the matrix $dF(p)$, $\eta_p^F = \|dF(p)\|$. The measure η_p^F satisfies three important properties: symmetry, coordinate system independence and stability. We now state and prove these properties.

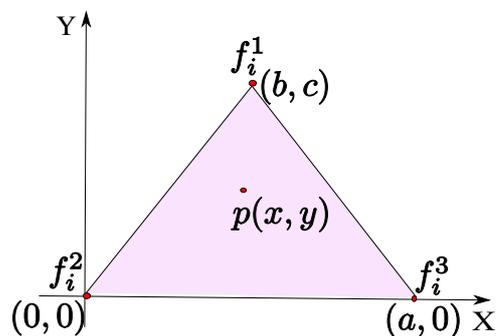


Figure 5.1: Piecewise-linear function defined on a triangle in a 2D mesh.

Symmetry

The measure is independent of the permutation of the functions in F . The proof follows directly from the definition.

Coordinate system independence

Let $(x'_1, x'_2, \dots, x'_n)$ denote a second orthonormal local coordinate system. Let J denote the Jacobian matrix that transforms the second coordinate system to the first. Since J represents a transformation between orthonormal coordinate systems, $|Jv| = |v|$ for $v \in \mathbb{R}^n$. This implies that $\|J\| = 1$. We first observe that if $dF'(p)$ is the derivative at p for the new coordinate system, then,

$$dF'(p) = dF(p)J.$$

This implies that $\|dF'(p)\| = \|dF(p)J\| \leq \|dF(p)\|\|J\|$ by applying Property 4 of the matrix norm. Since $\|J\| = 1$, we have $\|dF'(p)\| \leq \|dF(p)\|$. Similarly, we can prove that $\|dF(p)\| \leq \|dF'(p)\|$ by considering the Jacobian that transforms the first coordinate system to the second. This implies that the matrix norms are equal independent of the coordinate system.

Stability

We prove robustness of the measure when the scalar functions in the set F are piecewise-linear functions defined on a triangle mesh. We first observe that the derivative in the interior of a

triangle is well defined and is a constant. To simplify the description, we prove stability for a 2D triangle mesh representing a surface. Extension to higher dimensions is straightforward.

Consider a triangle in the mesh with coordinates as shown in Figure 5.1. The function value at any interior point $p(x, y)$ is given by

$$f_i^p = (A_1 f_i^1 + A_2 f_i^2 + A_3 f_i^3)x + (B_1 f_i^1 + B_2 f_i^2 + B_3 f_i^3)y,$$

where the constants $(A_1, A_2, A_3, B_1, B_2, B_3)$ depend only on a, b , and c . Consider a perturbation $\bar{F} = \{\bar{f}_1, \dots, \bar{f}_m\}$ of the functions in the set F , where $\bar{f}_i = f_i + \varepsilon_i$ and ε_i assumes small values. The partial derivatives satisfy the following relationship:

$$\frac{\partial \bar{f}_i}{\partial x} - \frac{\partial f_i}{\partial x} = (A_1 \varepsilon_i^1 + A_2 \varepsilon_i^2 + A_3 \varepsilon_i^3)$$

and

$$\frac{\partial \bar{f}_i}{\partial y} - \frac{\partial f_i}{\partial y} = (B_1 \varepsilon_i^1 + B_2 \varepsilon_i^2 + B_3 \varepsilon_i^3).$$

Therefore, the difference between the derivative matrices is

$$\begin{aligned} & d\bar{F}(p) - dF(p) \\ &= \begin{bmatrix} A_1 \varepsilon_1^1 + A_2 \varepsilon_1^2 + A_3 \varepsilon_1^3 & B_1 \varepsilon_1^1 + B_2 \varepsilon_1^2 + B_3 \varepsilon_1^3 \\ \vdots & \vdots \\ A_1 \varepsilon_m^1 + A_2 \varepsilon_m^2 + A_3 \varepsilon_m^3 & B_1 \varepsilon_m^1 + B_2 \varepsilon_m^2 + B_3 \varepsilon_m^3 \end{bmatrix} \\ &= \begin{bmatrix} \varepsilon_1^1 & \varepsilon_1^2 & \varepsilon_1^3 \\ \vdots & \vdots & \vdots \\ \varepsilon_m^1 & \varepsilon_m^2 & \varepsilon_m^3 \end{bmatrix} \begin{bmatrix} A_1 & B_1 \\ A_2 & B_2 \\ A_3 & B_3 \end{bmatrix}. \end{aligned}$$

Using Property 3 of the matrix norm,

$$|\eta_p^{\bar{F}} - \eta_p^F| \leq \|d\bar{F}(p) - dF(p)\|.$$

Let $\varepsilon = \sup_{1 \leq i \leq m, p \in \mathbb{M}} |\varepsilon_i(p)|$. Using Property 4 of the matrix norm to rewrite $\|d\bar{F}(p) - dF(p)\|$, we get

$$\begin{aligned}
|\eta_p^{\bar{F}} - \eta_p^F| &\leq \|d\bar{F}(p) - dF(p)\| \\
&\leq \left\| \begin{array}{ccc} \varepsilon_1^1 & \varepsilon_1^2 & \varepsilon_1^3 \\ \vdots & \vdots & \vdots \\ \varepsilon_m^1 & \varepsilon_m^2 & \varepsilon_m^3 \end{array} \right\| \left\| \begin{array}{cc} A_1 & B_1 \\ A_2 & B_2 \\ A_3 & B_3 \end{array} \right\| \\
&\leq \sqrt{\sum_{i=1}^m \sum_{j=1}^3 (\varepsilon_i^j)^2} \left\| \begin{array}{cc} A_1 & B_1 \\ A_2 & B_2 \\ A_3 & B_3 \end{array} \right\| \\
&\leq (\sqrt{3m})\varepsilon \left\| \begin{array}{cc} A_1 & B_1 \\ A_2 & B_2 \\ A_3 & B_3 \end{array} \right\|. \tag{5.1}
\end{aligned}$$

When the data is available over a structured grid and linearly interpolated along each coordinate axis, the difference between neighboring points in each axis direction can be used to approximate the partial derivatives at sample points and hence compute η^F . We show that the multifield comparison measure is stable when we use such an approximation. For simplicity, we assume that the domain is a 2-dimensional grid with each cell of size 1×1 units. Therefore, at grid point $p = (x, y)$,

$$\eta_p^F = \left\| \begin{array}{cc} f_1^{x+1,y} - f_1^{x,y} & f_1^{x,y+1} - f_1^{x,y} \\ \vdots & \vdots \\ f_m^{x+1,y} - f_m^{x,y} & f_m^{x,y+1} - f_m^{x,y} \end{array} \right\|.$$

It is easy to see that

$$\begin{aligned}
|\eta_p^{\bar{F}} - \eta_p^F| &\leq \left\| \begin{array}{cc} \varepsilon_1^{x+1,y} - \varepsilon_1^{x,y} & \varepsilon_1^{x,y+1} - \varepsilon_1^{x,y} \\ \vdots & \vdots \\ \varepsilon_m^{x+1,y} - \varepsilon_m^{x,y} & \varepsilon_m^{x,y+1} - \varepsilon_m^{x,y} \end{array} \right\| \\
&\leq \sqrt{\sum_{i=1}^m ((\varepsilon_i^{x+1,y} - \varepsilon_i^{x,y})^2 + (\varepsilon_i^{x,y+1} - \varepsilon_i^{x,y})^2)} \\
&\leq 2\sqrt{2m\varepsilon}
\end{aligned} \tag{5.2}$$

Equations 5.1 and 5.2 indicate that a finite change in the functions results in a bounded change in the multifield comparison measure. In the case of piecewise linear functions, the amount of change additionally depends on the size of the triangle.

The rows in the matrix $dF(p)$ represent the gradient vectors of the function. It is therefore easy to see (using property 3 from section 5.2.1) that a bounded change in the gradients will also result in a bounded change in the comparison measure. The multifield comparison measure is therefore robust with respect to perturbations in the scalar functions as well as their gradients. For smooth functions, the latter property still holds.

5.3 Analyzing Synthetic Functions

We describe the motivation for the definition of the multifield comparison measure by considering the case of one and two analytic functions. Next, we show how the comparison measure naturally extends to a larger number of scalar functions. We also show that the comparison measure can be used to capture the variation in gradient vectors over time for time varying scalar fields and to capture variation in time-varying vector fields.

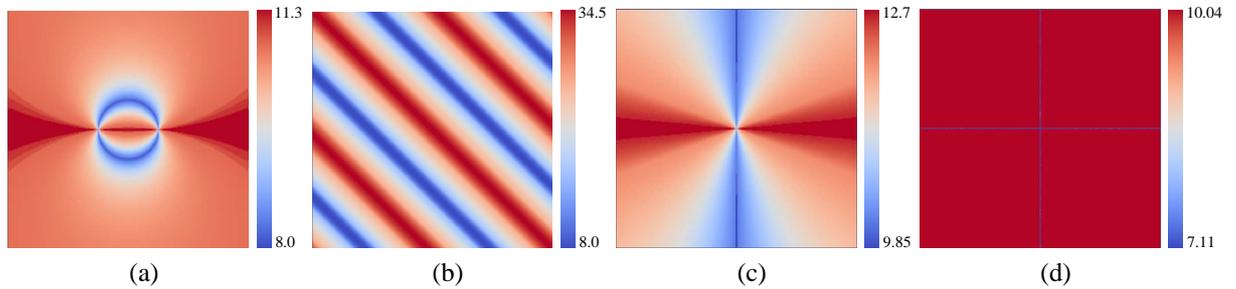


Figure 5.2: Multifield comparison measure η^F computed for synthetic functions defined on a 2D grid with the center as origin. (a) Two functions $f_1(x,y) = \sqrt{(x - 0.25)^2 + y^2}$ and $f_2(x,y) = \sqrt{(x + 0.25)^2 + y^2}$. The measure η^F attains high values on the Jacobi set and low values where the gradients are orthogonal. (b) The sinusoidal function $f_1(x,y) = \sin(3(x+y))$ and the linear function $f_2(x,y) = y$. (c) Three functions $f_1(x,y) = \sqrt{x^2 + y^2}$, $f_2(x,y) = \frac{1}{2}(\sqrt{3}x + y)$, and $f_3(x,y) = \frac{1}{2}(-\sqrt{3}x + y)$. (d) One hundred different scalar functions, whose gradient vectors have unit magnitude and directions are chosen uniformly at random at points on the two axes and are chosen to be some constant at remaining points on the plane.

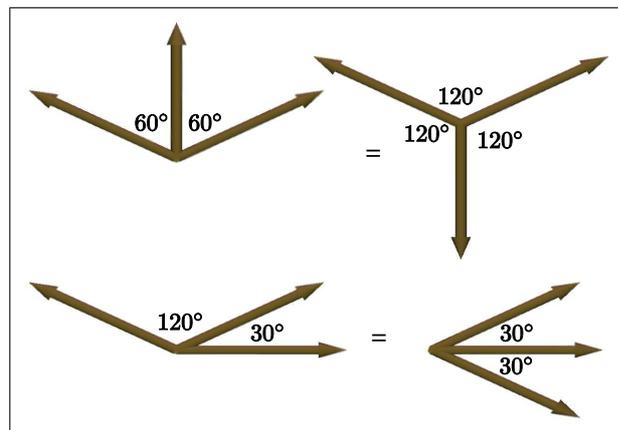


Figure 5.3: Two pairs of equivalent configurations of gradients of three functions described in Figure 5.2c. Gradient vectors subtend an angle of 120° at points along the Y-axis (top) and are more closely aligned with each other at points along the X-axis.

5.3.1 One / two scalar functions

In the special case of a single scalar function, the comparison measure at a point p is equal to the maximum rate of change of the function, which is essentially the length of the gradient vector at p . The gradient vector and its length have been used to study the behavior of a function with respect to the domain [30]. The proposed measure η^F is a natural extension of the notion of gradients and their relationship to multiple functions.

In the case of two smooth scalar functions, the gradients along with their mutual alignment is an indicator of the relationship between them [2, 3, 4]. Given two non-zero gradients, the multifield comparison measure η^F assumes the highest value when the gradients are parallel or anti-parallel. This set of points where the gradients align is called the Jacobi set [5]. The Jacobi set has been previously used to study the relationship between scalar functions [3]. The comparison measure assumes a minimum when the gradients are orthogonal. Orthogonality of the gradients indicates mutual independence of the functions. Figure 5.2a shows the computed comparison measure for two paraboloids, $f_1(x,y) = \sqrt{(x-0.25)^2 + y^2}$ and $f_2(x,y) = \sqrt{(x+0.25)^2 + y^2}$ defined on a 2D grid. The Jacobi set is the line $y = 0$ shown in dark red. The dark blue circle joining the centers of the paraboloids is the set of points where the gradients are orthogonal. Figure 5.2b depicts another example, a sinusoidal function $f_1(x,y) = \sin(3(x+y))$ and a linear function $f_2(x,y) = y$. The comparison measure assumes high values at the Jacobi set (shown in bright red) and the set of points with orthogonal alignment of gradients has low values (shown in blue). We note that when the gradient of a function is replaced with its negative, the measure remains the same. Two different configurations of the gradients yield the same comparison measure if it is possible to make a transition from one to another by replacing gradients with their negatives.

5.3.2 Multiple / time-varying scalar functions

Consider three functions $f_1(x,y) = \sqrt{x^2 + y^2}$, $f_2(x,y) = \frac{1}{2}(\sqrt{3}x + y)$, and $f_3(x,y) = \frac{1}{2}(-\sqrt{3}x + y)$. The multifield comparison measure (see Figure 5.2c) is minimum along the Y -axis. The

gradient vectors at points on the Y -axis subtend an angle of 120° with each other, whereas they are more aligned at points on the X -axis, see Figure 5.3. Figure 5.2d shows the comparison measure computed for hundred different scalar functions defined on a grid. All gradients have unit magnitude and the direction of the gradients for points on the X and Y axis are chosen uniformly at random. The gradients of all scalar functions are chosen to be equal at other points on the plane. We observe that the values of the comparison measure on the two axes are low compared to the values elsewhere on the grid. This indicates that given a set of gradient vectors with fixed magnitudes, the measure takes high values where the directions are more “coherent”.

Given a single time varying scalar field, we construct the set F of multiple scalar functions with one function corresponding to each time step. The multifield comparison measure in this case measures the variation of the scalar function over time. We extend the measure to compare multiple vector fields or analyze the variation in time-varying vector fields by replacing each row in the derivative matrix $dF(p)$ with the input vector at the point p .

5.4 Computation

Evaluating the multifield comparison measure at a point requires the solution to a maximization problem. In this section, we describe how this computation can be reduced to the faster evaluation of the maximum eigenvalue of a positive semi-definite matrix.

5.4.1 Maximum eigenvalue computation

From the definitions of the multifield comparison measure and the norm of a matrix, we have

$$\eta_p^F = \left(\max_{x \in \mathbb{R}^n, \|x\|=1} x^T (dF(p))^T (dF(p)) x \right)^{\frac{1}{2}}.$$

We rewrite the matrix product $(dF(p))^T(dF(p))$ as $U^T \Lambda U$, where U is an orthogonal matrix and Λ is a diagonal matrix consisting of the eigenvalues of $(dF(p))^T(dF(p))$ as entries in its diagonal. This follows from the spectral theorem from linear algebra [46]:

$$\eta_p^F = \left(\max_{x \in \mathbb{R}^n, \|x\|=1} x^T U^T \Lambda U x \right)^{\frac{1}{2}}.$$

Since the orthogonal matrix U represents a length preserving and invertible transformation, we can write the above expression as

$$\begin{aligned} \eta_p^F &= \left(\max_{x \in \mathbb{R}^n, \|x\|=1} x^T \Lambda x \right)^{\frac{1}{2}} \\ &= \max\{\sqrt{\lambda} : \lambda \text{ is a diagonal element of } \Lambda\} \\ &= \max\{\sqrt{\lambda} : \lambda \text{ is an eigenvalue of } (dF(p))^T(dF(p))\}. \end{aligned}$$

For piecewise linear functions defined on a triangle mesh, the derivative matrix $dF(p)$ is constant within a triangle and can be computed by choosing a local coordinate system.

5.4.2 Analysis

The size of the $n \times n$ matrix $(dF(p))^T(dF(p))$ depends only on the dimension of the domain. Therefore, the time taken for computing the eigen values of the product matrix also depends only on the dimension of the domain and is, in particular, independent of the number of fields m .

5.5 Applications

We use the multifield comparison measure to study various real-world data from weather modeling, climate simulations, and combustion simulations. Observations on the combustion data were compared with prior work described in the literature. First, we study a simulation of the

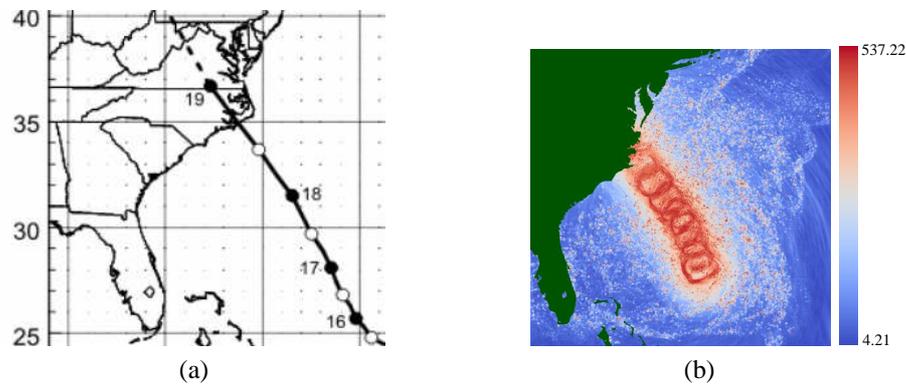


Figure 5.4: (a) The hurricane track released by the US National Hurricane Center (source: <http://www.nhc.noaa.gov>). The track relevant to the period of simulation is between point 17 and point 19 when the hurricane struck the coast. (b) Multifield comparison measure computed for nine pressure fields. The region in red with high values of the comparison measure corresponds to the trace of the eye of the hurricane. Land is shown in green.

hurricane Isabel. Next, we apply our multifield comparison measure on a global wind pattern data set. Finally, we study a combustion simulation data set by aggregating the multifield comparison measure over the domain at each time step. The data in the following applications varies in the dimensionality of the domain and the number of fields – two scalar fields defined on a 3D domain, multiple time-varying scalar fields defined on 2D and 3D domains, and time-varying vector fields. In all cases the analysis is based on the proposed multifield comparison measure η^F . We perform experiments on data sets obtained using a variety of climate models. In Section 5.5.2, we describe results from four different climate models, which demonstrate the applicability of the proposed method.

5.5.1 Isabel Hurricane

Hurricane Isabel struck the west Atlantic region in September 2003. A simulation of the phenomenon was performed on a $600 \times 600 \times 100$ grid corresponding to a physical volume of $2139\text{km} \times 2004\text{km} \times 19.8 \text{ km}$ over 48 time steps corresponding to 48 simulated hours [33]¹.

¹Hurricane Isabel data was produced by the Weather Research and Forecast (WRF) model, courtesy of NCAR and the U.S. National Science Foundation (NSF).

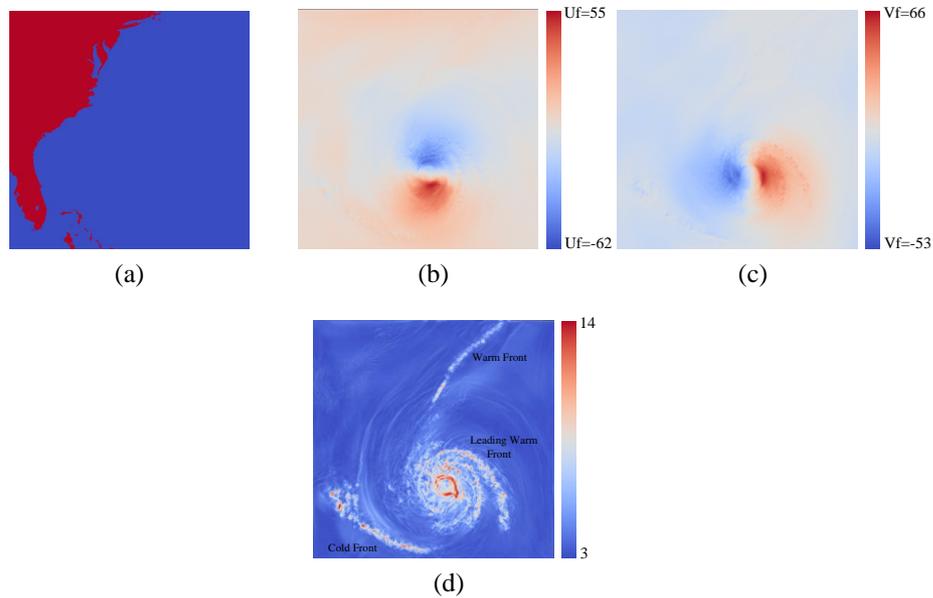


Figure 5.5: Fronts in Hurricane Isabel at hour 10. (a) Region of simulation. Land mass is shown in red. (b) Volume rendering (top view) of horizontal wind speed U_f . (c) Volume rendering (top view) of horizontal wind speed V_f . (d) Volume rendering (top view) of multifield comparison measure η^F computed for U_f and V_f showing the rainbands at different fronts. The location of the fronts is not available from the individual scalar fields U_f and V_f .

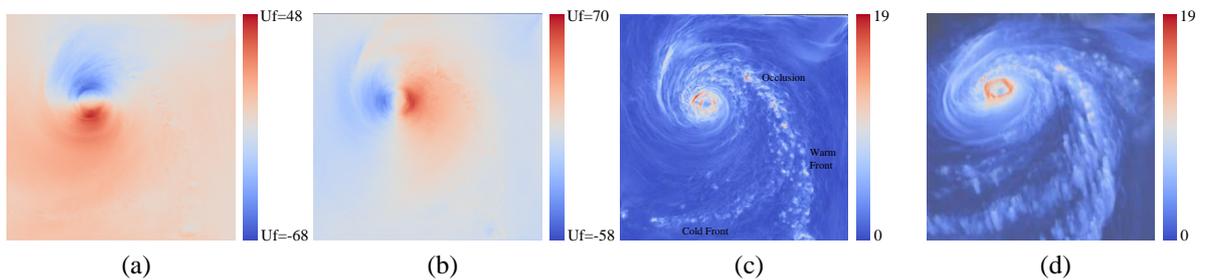


Figure 5.6: Fronts in Hurricane Isabel at hour 40. (a) Volume rendering (top view) of horizontal wind speed U_f . (b) Volume rendering (top view) of horizontal wind speed V_f . (c) Volume rendering (top view) of multifield comparison measure η^F computed for U_f and V_f showing the rainbands at different fronts. The cold front leads the warm front resulting in an occlusion. (d) Volume rendering from a different viewpoint .

Several quantities were computed by the simulation. The scalar fields relevant to our experiment are pressure (Pf) and the horizontal wind velocity components (Uf and Vf).

Hurricane track

The horizontal pressure gradient can be observed to be high near the eye of the hurricane throughout the simulation. We study the pressure field at every fifth time step defined on a 2D slice corresponding to altitude 1500m. Figure 5.4b shows the multifield comparison measure computed for the nine pressure fields. The red circular regions that correspond to high values of the comparison measure correspond to the hurricane track. Figure 5.4a shows the track provided by the US National Hurricane Center for reference. The results are similar when we compute the multifield comparison measure for all 48 pressure fields. One of the horizontal pressure gradients assumes a high value at the eye compared to the rest and hence the comparison measure is higher compared to regions far from the path of the eye.

Rainbands and front

Cloud structures associated with an area of rainfall, called rainbands, occur mainly at boundaries separating two masses of air of different densities and temperatures, called fronts. The leading edge of the cooler mass of air is called the cold front and the leading edge of a warm air mass is called the warm front. The turbulence of the horizontal wind velocity is high near rain bands. We study the fronts by computing the multifield comparison measure for the pair of 3D scalar fields Uf and Vf, where the 3D domain corresponds to the volume in the altitude range 1500m-5800m.

First, we compute the multifield comparison measure for the fields Uf and Vf in the 10th time step. Figure 5.5 shows the result of our experiment as volume rendered images with the view point located above the volume. In particular, Figure 5.5d shows the location of two warm fronts and a cold front. This information about fronts cannot be extracted from the two functions individually (see Figures 5.5b and 5.5c). The comparison measure successfully

captures the relationship between the fields at the fronts. The warm front leads the cold front in the 10th time step. The precipitation structures indicated by light red spots on the rain bands on the land mass were responsible for heavy rainfall in Florida. Next, we compute the multifield comparison measure for the fields U_f and V_f in the 40th hour of simulation. We observe that the warm front at the north has disappeared, see Figures 5.6c and 5.6d . The previously leading warm front is overtaken by the cold front resulting in an occlusion.

Value of Study

Both structures discussed above are key to a comprehensive study of the hurricane. The track of a hurricane or a cyclone generated from a forecast is helpful in predicting the areas susceptible to severe weather. Fronts often give valuable information about severe weather to the forecaster. Rainbands at cold fronts are often strong in nature and can be responsible for heavy thunder storms. Typically, occlusion fronts are associated with thunder storms and their passage results in the reduction of humidity.

5.5.2 Global Wind Patterns

Prevailing winds are winds that blow in a dominant direction at a particular point. Movements in the Earth's atmosphere affect these winds. In regions of mid-latitudes, the winds blow from west to the east and are known as westerlies. The winds found in the tropics near the equator are easterlies or trade winds. Figure 5.7a shows the different prevailing winds on earth. We study wind patterns on earth using a climate simulation of 50 years between 1960 and 2009 [47]². The data is available for 600 time steps corresponding to each month over the period of simulation. Each time step is a 3D grid with resolution corresponding to $1^\circ \times$

²The climate data was part of the WCRP CMIP3 Multi-Model data repository at <https://esgcat.llnl.gov:8443>. We acknowledge the modeling groups for making their model output available for analysis, the Program for Climate Model Diagnosis and Intercomparison (PCMDI) for collecting and archiving this data, and the WCRP's Working Group on Coupled Modelling (WGCM) for organizing the model data analysis activity. The WCRP CMIP3 multi-model dataset is supported by the Office of Science, U.S. Department of Energy.

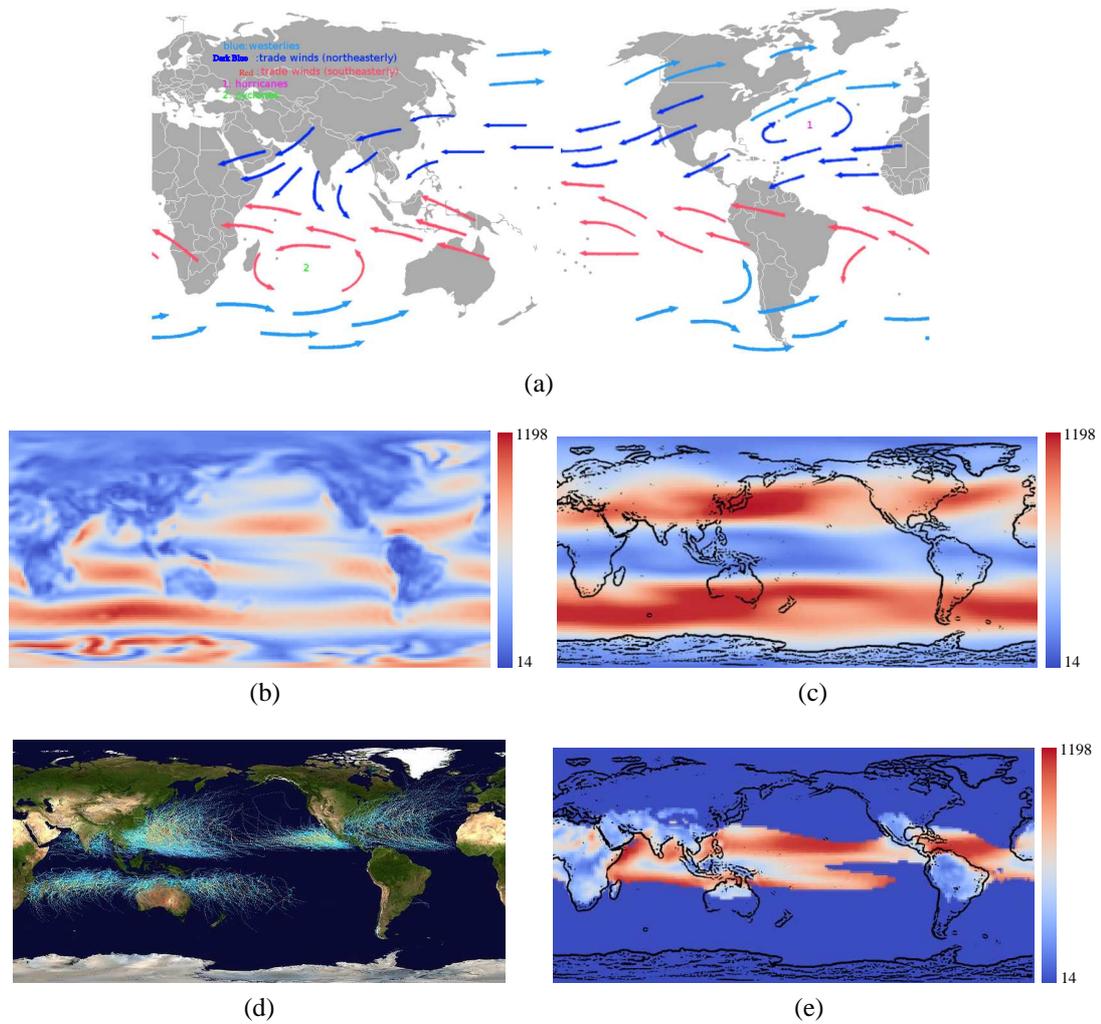


Figure 5.7: Multifield comparison measure η^F computed for wind velocities over the years 1960-2009, where the comparison is over a set of six hundred 3D vector fields. (a) Map of world showing wind patterns (source: Wikipedia) (b) Distribution of η^F over surface corresponding to pressure elevation 925 hPa. The dark red regions correspond to the wind patterns. (c) Distribution of η^F over surface corresponding to pressure elevation 300 hPa. The temperate regions exhibit higher values. (d) Storm track for the years 1985-2005 (source: Wikipedia) (e) Distribution of η^F after removing regions with low mean temperature ($< 27^\circ\text{C}$). Red regions correspond to the storm tracks. The world map is overlaid for clarity.

$1^\circ \times 16plev$ (pressure elevations) on earth. Pressure elevations correspond to pressures varying from 1000 hPa on the surface to 30 hPa in the upper atmosphere.

Westerlies and trade wind

The wind velocity at a point on the grid is a vector quantity. We use the monthly wind velocities as rows in the derivative matrix and compute the matrix norm for 600 vector fields. The norm η^F measures the variation of the wind velocities over a time period of 50 years. Figure 5.7b shows the distribution of the computed comparison measure over a surface corresponding to pressure elevation 925 hPa. Comparing with wind patterns in Figure 5.7a, we see that the measure assumes high values in regions that lie in the path of prevailing winds, particularly the westerlies found in the regions surrounding Antarctica, the region of hurricanes in Atlantic, the cyclone prone region between Madagascar and Australia, and the trade winds across the Atlantic sea traveling towards the Caribbean sea. The distribution of the comparison measure over the isobar for pressure level 300 hPa, which corresponds to approximately 30000 feet above sea level, is shown in Figure 5.7c. The values of the comparison measure are higher compared to Figure 5.7b because friction and other effects can cause the wind flow at 925 hPa to be less steady than at higher levels such as 300 hPa. We note that the comparison measure assumes high values over the temperate regions corresponding to the westerly jet. This is a semi-permanent feature of the mid-latitudes. Many regions in the tropics undergo a seasonal reversal of wind (called the monsoons). Lower values of the comparison measure over the tropics indicates unsteadiness and corresponds to a seasonal reversal in wind pattern over this part of the world.

Storm track

The regions over the ocean with warm temperatures ($> 27^\circ\text{C}$) are susceptible to storms. We filter out regions with lower temperatures and restrict our analysis to the months from June to November with the aim of locating storm tracks. Regions shown in blue in Figure 5.7e

have been filtered out. The red regions match closely with the storm tracks shown in Figure 5.7d. We notice that even though the west coast of South America has trade winds, storms are particularly absent due to lower temperatures. The storm prevalent regions in the Indian, Atlantic, and Pacific oceans have high values of the comparison measure. To ensure the validity of our results, we compute wind patterns for other models: cccma-cgcm3 [48, 49] and bccr-bcm2.0 [50]. (see Figure 5.9)

We next study the changes in storm tracks over two centuries. For each century, the storm track is computed over a period of 25 years (see Figure 5.8). The distribution of the difference between both images indicate that there could be changes in storm patterns with possibly stronger winds over the Indian Ocean, near Phillipines, the Central Pacific and off the coast of Indonesia/Australia. The conditions could be less stormy off the American coasts (both east and west), northern and southern Pacific and off the Chinese coast. This however does not preclude individual storms being of higher strength as suggested by Webster et al. [51]. Also, these results could be dependent upon the ECHAM5 coupled ocean-atmosphere model used. We note that the signals are stronger in the figure on the right.

Value of Study

Wind patterns give the details of the wind over a particular place. They cause various local and global phenomena and are widely studied by climate scientists. For example, the trade winds are responsible for tropical cyclones over oceans. Storm tracks generated using winds give us information on regions where storms are more probable. Computing storm tracks for long periods would require ascertaining tracks of every individual hurricane and cyclone and plotting them. Our approach simplifies this computation by considering all 600 time steps together to generate the distribution of the comparison measure.

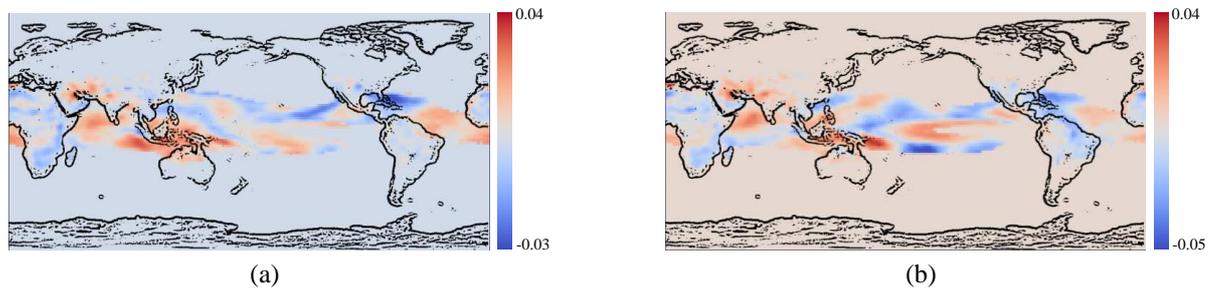


Figure 5.8: Changes in storm tracks studied by computing the difference in multifield comparison measure computed for two periods of 25 years. (a) η^F for the years 1985-2010 subtracted from η^F for the years 1885-1909. (b) η^F for the years 2075-2099 subtracted from η^F for the years 1885-1909.

5.5.3 Hydrogen Combustion

We study phases in the combustion of an inhomogeneous turbulent mixture of fuel and oxidizer. Hydrogen fuel at 300K is mixed with an oxidizer (21% oxygen) at 1200K. The influence of turbulence on the different phases of combustion is studied in the simulation³. The compressed fuel ignites at multiple spots because of the inhomogeneity in the air-fuel ratio. Depending on the air-fuel ratio, the flame either propagates in an outward direction from the ignition spot or burns out. Further details of the computation can be found in the description of the simulation by Echehki and Chen [19] and in the description of a visual analysis of this data by Koeqler [32]. The domain of the simulation is a 600×600 grid for 67 time steps. The species mass fractions of the fuel H_2 , oxygen O_2 , and intermediate HO_2 are given at each grid point for all time steps. The reactions between these different radicals determine the phases of combustion. We study these phases with the multifield comparison measure.

Phases of combustion

We aggregate the multifield comparison measure for a time step t by computing its integral over the domain D of simulation:

$$\eta^{F,t} = \int_{x \in D} \eta_x^F dx.$$

³We would like to thank Valerio Pascucci and Jackie Chen for providing the combustion data

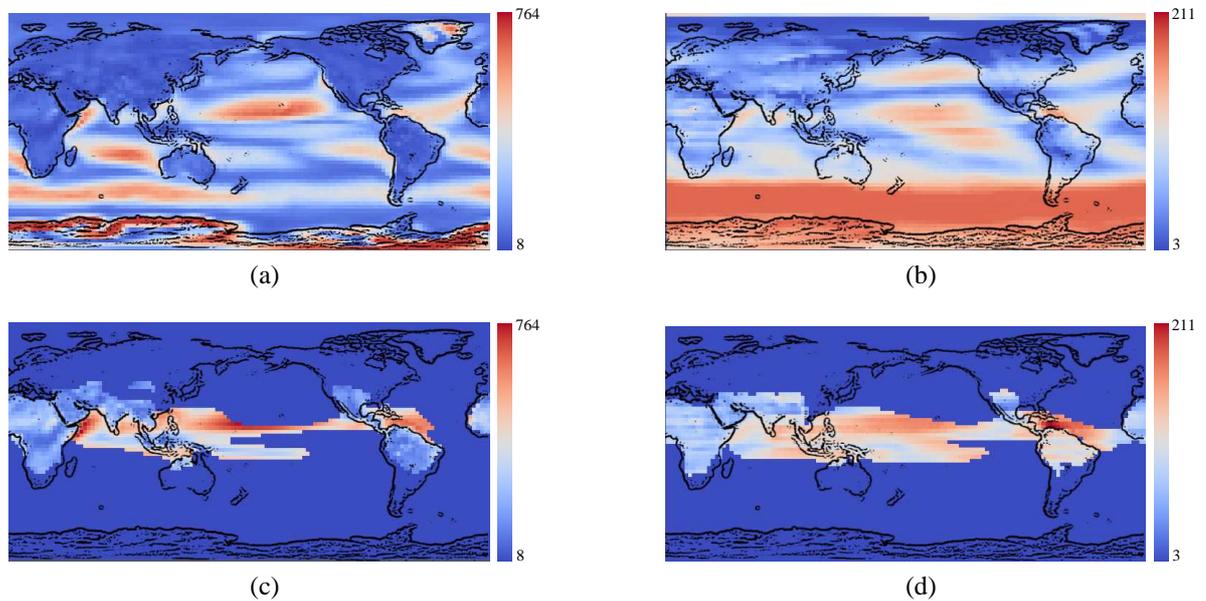


Figure 5.9: Study of wind patterns and storm tracks using data from different climate models. Global wind pattern captured by the multifield comparison measure: (a) model bccr-bcm2.0 and (b) model cccma-cgcm3. Storm track for the years 1976-2000 highlighted by the multifield comparison measure: (c) model bccr-bcm2.0 and (d) model:cccma-cgcm3.

In our study, we consider all three scalar fields for computing the comparison measure for a time step, i.e. $F = \{H_2, O_2, HO_2\}$. The plot of $\eta^{F,t}$ over time is shown in Figure 5.10. We compare our results with a plot of the global comparison measure κ , introduced by Edelsbrunner et al. [3], see Figure 5.10. The fields used for computing κ are H_2 and O_2 . The maximum number of fields that κ can compare over a 2D domain is two, whereas we are able to consider all three fields in our analysis.

Following the work of Koegler, the areas in the domain that eventually ignite can be considered as features. In the pre-ignition phase, the concentrations of the intermediate radicals build up in regions that have sufficient mass fraction of H_2 . The number of features attains a maximum during time steps 7-14 [32]. This is captured by a knee in the plot of $\eta^{F,t}$. The plot of κ does not indicate changes that happen in the pre-ignition build up of radicals because the intermediate HO_2 is not considered in the computation. Ignition (time step 28) happens at areas with high radical concentrations during the ignition phase and the flame front spreads to hot enough areas with the right mix of fuel and oxidizer during the burning phase (time steps

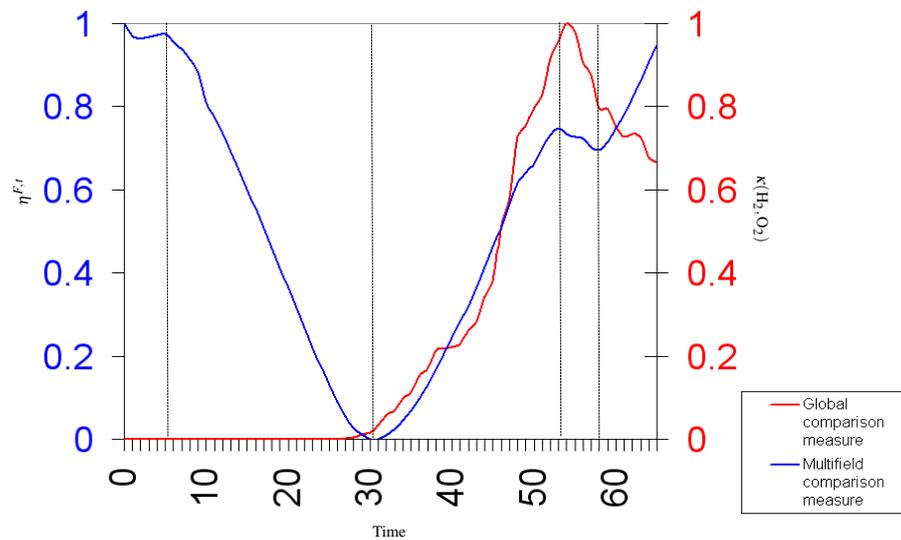


Figure 5.10: Analyzing phases of combustion using an aggregate $\eta^{F,t}$ of the multifield comparison measure over the domain within a time step t . The set $F = \{\text{H}_2, \text{O}_2, \text{HO}_2\}$. The plot of $\eta^{F,t}$ (blue) over time captures more phases of the combustion process compared to the plot of the global comparison measure $\kappa(\text{H}_2, \text{O}_2)$ (red) proposed by Edelsbrunner et al. [3]. The vertical dashed lines approximately correspond to the four phases of combustion: pre-ignition, ignition, burning, and extinction.

50-55). Ignition and burning are captured by a minimum and a maximum respectively in the plot of $\eta^{F,t}$. These phases are also captured by κ because of the interaction between H_2 and O_2 during this time period. The beginning of the extinction phase (approximately time step 60) where the flame begins to extinguish is also captured more clearly by a minimum in the plot of $\eta^{F,t}$. We believe that the reason $\eta^{F,t}$ is able to capture more information compared to κ is because all three fields play a role in defining the phases.

5.6 Discussion

We now discuss some limitations of the multifield comparison measure and describe an experiment to study the sensitivity of the measure to noise in the input. We also discuss its relationship with principal component analysis (PCA).

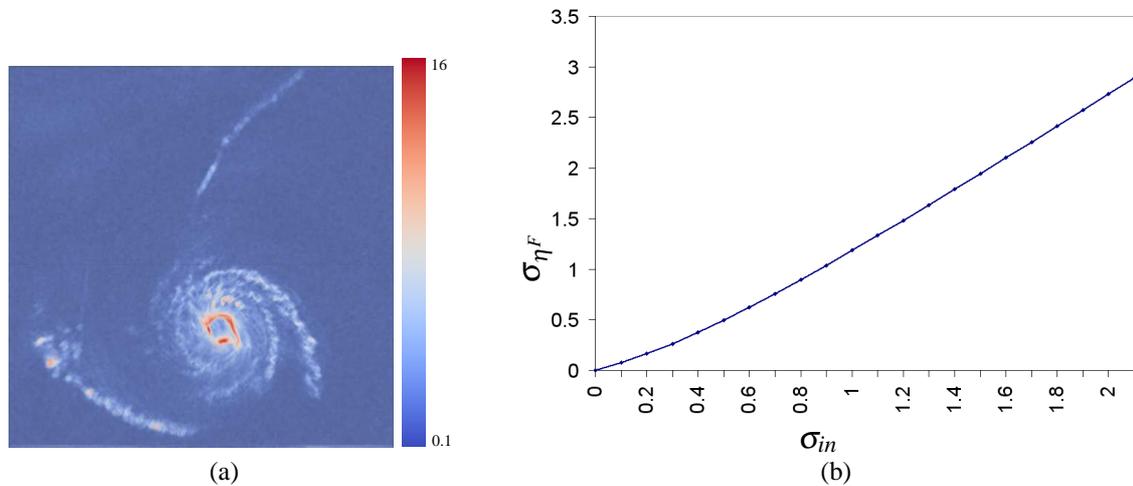


Figure 5.11: Stability in the comparison measure for Isabel. (a) Volume rendering (top view) of multifield comparison measure η^F computed after adding Gaussian noise (standard deviation = 1) to the fields U_f and V_f . The rain bands are still clearly visible. (b) Graph showing near linear relationship between the standard deviation of the noise in the input σ_{in} , and the mean deviation of the comparison measure σ_{η^F} .

5.6.1 Limitations

The basic premise behind using the comparison measure to capture relationships in multifield real world data is the fact that the agreement among the different gradient fields can often capture interactions among fields. In many applications, where this does not hold good, using the comparison measure or other gradient based comparison measures would be less fruitful. We therefore believe that gradient based techniques such as ours complements other well known techniques like the local statistical complexity [43] and the Pearson correlation coefficient. Another drawback of the comparison measure is its sensitivity to scaling of individual fields.

5.6.2 Sensitivity to noise

We validate our claim that the comparison measure is insensitive to noise using the hurricane data described in Section 5.5.1. The different features present in the data like the weather fronts are still clearly visible in the computed comparison measure field after adding a Gaussian noise (standard deviation = 1) to the input fields U_f and V_f of the 10th time step in the

simulation, see Figure 5.11a. It should be noted that the noise we have added is large and quite unlikely to occur in real data sets. We also study the mean deviation in the comparison measure computed on input with Gaussian noise of different amplitudes. Figure 5.11b shows the linear relationships between the observed mean deviation of η^F and the noise, which indicates that the multifield comparison measure is not much sensitive to noise.

5.6.3 Multifield comparison measure and PCA

In Section 5.4, we showed that the multifield comparison measure can be computed by finding the maximum eigenvalue of the matrix $dF^T dF$. If the components of the gradients have a zero mean, the direction of the eigenvector corresponding to the maximum eigenvalue is the direction in which the variance of the inner products of the gradients with the eigenvector is maximum. The multifield comparison measure is therefore the variance of the principal component when performing principal component analysis on the matrix dF [52]. The matrix $dF^T dF$ can be considered the covariance matrix.

5.7 Conclusions

We have described a robust multifield comparison measure for scalar fields whose distribution over the domain plays an important role in the visual analysis of the input fields. The comparison measure is computed locally at all points of the domain as the maximum eigenvalue of a small sized matrix. We described applications of the comparison measure to study various simulation datasets from climate science and combustion studies where the data is represented using multiple 2D, 3D, or time-varying scalar fields. We used the comparison measure to study up to 600 scalar fields defined on the domain. We list three ideas for future work:

- The multifield comparison measure being sensitive to the scaling of individual fields may not be always desirable. One approach to address this issue is to scale the scalar fields or normalize their gradients. Though this method would work in some cases, it

could bias the results by scaling up small insignificant gradients. A complete solution remains to be a challenge.

- Integrating the multifield comparison measure into the query-based visualization framework of Gosink et al. [2] will be an interesting task.
- Identification of important isovalues of a scalar function in multifield data is a challenging problem (see Section 4.1 in Chapter 4). A global version of the proposed comparison measure may help locate these isovalues.

Chapter 6

Conclusions

In this thesis, we have argued that to effectively understand and visualize multifield scientific data, it is necessary to consider the interactions that exist between the different fields. We have restricted our focus to scalar fields defined on manifolds. Whenever the field is given as samples over a simplicial complex, we have used piecewise linear interpolation to reconstruct the original field.

We have described three new techniques to understand and visualize inter-variable relationships in multifield scientific data. For each technique, we have described real world applications where the technique could be used. Also, we have discussed future work and limitations at the end of the corresponding chapter. We summarize the thesis contributions in the following list :

- We have introduced a new technique to compute and simplify the Jacobi set of two morse functions. Our approach allows the representation of the Jacobi set at multiple levels of simplification.
- We have developed software to compute, simplify and view the simplified Jacobi set.
- We have introduced a new variation density function to identify interesting isosurfaces in multifield data.

- We developed an efficient parallel implementation to identify interesting isovalues using the variation density function.
- We have introduced a new multifield comparison measure that captures relationships between an arbitrary number of scalar fields. We also have an efficient implementation to compute the multifield comparison measure.

Throughout the thesis, we have assumed that the gradients along with their mutual alignment play a significant role in defining relationships between fields. Extending the proposed techniques to data where this assumption is false is a challenging problem. We believe that our techniques complement existing techniques for multifield visualization.

Bibliography

- [1] M. Ward, G. Grinstein, and D. Keim, *Interactive Data Visualization, Foundations, Techniques, and Applications*. A K Peters Ltd, 2010.
- [2] L. Gosink, W. Anderson, J. Bethel, and K. Joy, “Variable interactions in query-driven visualization,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 6, pp. 1400–1407, 2007.
- [3] H. Edelsbrunner, J. Harer, V. Natarajan, and V. Pascucci, “Local and global comparison of continuous functions,” in *Proc. IEEE Conf. Visualization*, 2004, pp. 275–280.
- [4] N. Sauber, H. Theisel, and H.-P. Seidel, “Multifield-graphs: An approach to visualizing correlations in multifield scalar data,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 12, no. 5, pp. 917–924, 2006.
- [5] H. Edelsbrunner and J. Harer, “Jacobi set of multiple morse funtions,” in *Foundations of Computational Mathematics, Minneapolis*. Cambridge Univ. Press, 2002, pp. 37–57.
- [6] J. Kniss, G. Kindlmann, and C. Hansen, “Interactive volume rendering using multi-dimensional transfer functions and direct manipulation widgets,” in *Proc. IEEE Conf. Visualization*, ser. VIS '01. Washington, DC, USA: IEEE Computer Society, 2001, pp. 255–262. [Online]. Available: <http://dl.acm.org/citation.cfm?id=601671.601711>
- [7] H. Carr, B. Duffy, and B. Denby, “On histograms and isosurface statistics,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 12, no. 5, pp. 1259–1266, 2006.

- [8] C. Bajaj, V. Pascucci, and D. Schikore, “The contour spectrum,” in *Proc. IEEE Conf. Visualization 1997*, Oct. 1997, pp. 167–173.
- [9] M. Golubitsky and V. Guillemin, *Stable Mappings and Their Singularities*. Springer, revised edition, 1974.
- [10] K. Engel, M. Hadwiger, J. M. Kniss, C. Rezk-Salama, and D. Weiskopf, *Real-Time Volume Graphics*. A K Peters Ltd, 2006.
- [11] F. Lekien and J. Marsden, “Tricubic interpolation in three dimensions,” *Journal of Numerical Methods and Engineering*, vol. 63, pp. 455–471, 2005.
- [12] W. E. Lorensen and H. E. Cline, “Marching cubes: A high resolution 3d surface construction algorithm,” *Computer Graphics*, vol. 21:4, 1987.
- [13] G. M. Treece, R. W. Prager, and A. H. Gee, “Regularised marching tetrahedra: Improved iso-surface extraction,” *Computers and Graphics*, vol. 23, pp. 583–598, 1998.
- [14] H. Edelsbrunner, D. Letscher, and A. Zomorodian, “Topological persistence and simplification,” *Discrete Comput. Geom.*, vol. 28, no. 4, pp. 511–533, 2002.
- [15] J. Bennett, V. Pascucci, and K. Joy, “Genus oblivious cross parameterization: Robust topological management of inter-surface maps,” in *Proceedings of the 15th Pacific Conference on Computer Graphics and Applications*. Washington, DC, USA: IEEE Computer Society, 2007, pp. 238–247.
- [16] P.-T. Bremer, E. M. Bringa, M. A. Duchaineau, A. G. Gyulassy, D. Laney, A. Mascarenhas, and V. Pascucci, “Topological feature extraction and tracking,” *Journal of Physics: Conference Series*, vol. 78, p. 012007 (5pp), 2007. [Online]. Available: <http://stacks.iop.org/1742-6596/78/012007>
- [17] H. Edelsbrunner, D. Morozov, and V. Pascucci, “Persistence-sensitive simplification

- functions on 2-manifolds,” in *Proc. Symposium on Computational geometry*, 2006, pp. 127–134.
- [18] M. Meyer, M. Desbrun, P. Schroder, and A. Barr, “Discrete differential geometry operators for triangulated 2-manifolds,” *VisMath.*, 2002.
- [19] T. Echehki and J. H. Chen, “Direct numerical simulation of autoignition in nonhomogeneous hydrogen-air mixtures,” *Combustion and Flame*, vol. 134, no. 3, pp. 169–191, 2003.
- [20] Y. Matsumoto, *An Introduction to Morse Theory*. Amer. Math. Soc., 2002, translated from Japanese by K. Hudson and M. Saito.
- [21] G. Reeb, “Sur les points singuliers d’une forme de pfaff complètement intégrable ou d’une fonction numérique.” *Comptes Rendus de L’Académie des Séances*, vol. 222, pp. 847–849, 1946.
- [22] V. Pekar, R. Wiemker, and D. Hempel, “Fast detection of meaningful isosurfaces for volume data visualization,” in *Proc. IEEE Conf. Visualization*, 2001, pp. 223–230.
- [23] S. Tenginakai, J. Lee, and R. Machiraju, “Salient iso-surface detection with model-independent statistical signatures,” in *Proc. IEEE Conf. Visualization*, 2001, pp. 231–238.
- [24] C. E. Scheidegger, J. M. Schreiner, B. Duffy, H. Carr, and C. T. Silva, “Revisiting histograms and isosurface statistics,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 14, no. 6, pp. 1659–1666, 2008.
- [25] H. Carr, J. Snoeyink, and U. Axen, “Computing contour trees in all dimensions,” in *Proc. ACM-SIAM Symposium on Discrete algorithms*, 2000, pp. 918–926.
- [26] V. Pascucci, G. Scorzelli, P.-T. Bremer, and A. Mascarenhas, “Robust on-line computation of reeb graphs: simplicity and speed,” *ACM Trans. Graph.*, vol. 26, no. 3, pp. 58.1–58.9, 2007.

- [27] H. Carr, J. Snoeyink, and M. van de Panne, "Simplifying flexible isosurfaces using local geometric measures," in *Proc. IEEE Conf. Visualization*, 2004, pp. 497–504.
- [28] P. K. Agarwal, H. Edelsbrunner, J. Harer, and Y. Wang, "Extreme elevation on a 2-manifold," *Discrete Comput. Geom.*, vol. 36, no. 4, pp. 553–572, 2006.
- [29] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer, "Extending persistence using Poincaré and Lefschetz duality," *Found. Comput. Math.*, vol. 9, no. 1, pp. 79–103, 2009.
- [30] A. Gyulassy, V. Natarajan, V. Pascucci, P.-T. Bremer, and B. Hamann, "A topological approach to simplification of three-dimensional scalar functions," *IEEE Transactions on Visualization and Computer Graphics*, vol. 12, no. 4, pp. 474–484, 2006.
- [31] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer, "Stability of persistence diagrams," *Discrete Comput. Geom.*, vol. 37, no. 1, pp. 103–120, 2007.
- [32] W. S. Koegler, "Case study: application of feature tracking to analysis of autoignition simulation data," in *Proc. IEEE Conf. Visualization*, 2001, pp. 461–464.
- [33] W. Wang, C. Bruyere, and B. Kuo, "Competition data set and description in 2004 IEEE Visualization design contest," 2004, <http://vis.computer.org/vis2004contest/data.html>.
- [34] D. Whalen and M. L. Norman, "Competition data set and description in 2008 IEEE Visualization design contest," 2008. [Online]. Available: <http://vis.computer.org/VisWeek2008/vis/contests.html>
- [35] A. Rajwade, A. Banerjee, and A. Rangarajan, "Probability density estimation using isocontours and isosurfaces: Applications to information-theoretic image registration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 3, pp. 475–491, 2009.
- [36] S. Bachthaler and D. Weiskopf, "Continuous scatterplots," *IEEE Transactions on Visualization and Computer Graphics*, vol. 14, no. 6, pp. 1428–1435, 2008.

- [37] C. Hansen and C. Johnson, *Visualization Handbook*. Academic Press, 2004.
- [38] W. Schroeder, K. Martin, and B. Lorensen, *The Visualization Toolkit An Object-Oriented Approach To 3D Graphics, 4th Edition*. Kitware, Inc., 2006.
- [39] W. Cai and G. Sakas, “Data intermixing and multi-volume rendering,” *Computer Graphics Forum*, vol. 18, no. 3, pp. 359–368, 1999.
- [40] C. Bajaj, V. Pascucci, G. Rabbio, and D. Schikore, “Hypervolume visualization: a challenge in simplicity,” in *Proc. IEEE Symp. Volume Visualization*, 1998, pp. 95–102.
- [41] J. Woodring and H.-W. Shen, “Multi-variate, time varying, and comparative visualization with contextual cues,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 12, no. 5, pp. 909–916, 2006.
- [42] T.-Y. Lee and H.-W. Shen, “Visualization and exploration of temporal trend relationships in multivariate time-varying data,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, pp. 1359–1366, 2009.
- [43] H. Jänicke, A. Wiebel, G. Scheuermann, and W. Kollmann, “Multifield visualization using local statistical complexity,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, pp. 1384–1391, 2007.
- [44] H. Jänicke, M. Bottinger, X. Tricoche, and G. Scheuermann, “Automatic detection and visualization of distinctive structures in 3d unsteady multi-fields,” *Computer Graphics Forum*, vol. 27, no. 3, pp. 767–774, 2008.
- [45] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.
- [46] K. Hoffman and R. Kunze, *Linear Algebra*. Prentice Hall, 1971.
- [47] E. Roeckner, G. Bäuml, L. Bonaventura, R. Brokopf, M. Esch, M. Giorgetta, S. Hagemann, I. Kirchner, L. Kornbluh, E. Manzini, A. Rhodin, U. Schlese, U. Schulzweida,

- and A. Tompkins, “The atmospheric general circulation model echam5. part 1: Model description,” Max Planck Institute of Meteorology, Tech. Rep. 349, 2003.
- [48] G. Boer, N. McFarlane, R. Laprise, H. J.D., and J. Blanchet, “The canadian climate centre spectral atmospheric general circulation model,” *Atmos. Ocean.*, vol. 22, pp. 397–429, 1984.
- [49] G. Flato and G. Boer, “Warming asymmetry in climate change,” *Simulations. Geophys. Res. Lett.*, vol. 28, pp. 195–198, 2001.
- [50] T. Furevik, M. Bentsen, H. Drange, I. K. T. Kindem, N. G. Kvamstø, and A. Sorteberg, “Description and evaluation of the bergen climate model: Arpege coupled with micom,” *Climate Dynamics*, vol. 21, pp. 27–51, 2003, 10.1007/s00382-003-0317-5. [Online]. Available: <http://dx.doi.org/10.1007/s00382-003-0317-5>
- [51] P. Webster, G. Holland, J. Curry, and H.-R.Chang, “Changes in tropical cyclone number, duration, and intensity in a warming environment,” *Science*, September 2005.
- [52] I. Jolliffe, *Principal Component Analysis*, 2nd ed. Springer Series in Statistics, 2002.

Index

- Comparison measure, 17, 37
- Critical point, 10, 13
- Eigen value, 68
- Gradient vector, 10
- Hessian, 15
- Homeomorphic spaces, 9
- Homeomorphism, 9
- Isabel hurricane, 51
- isocontour, 12
- isoline, 12
- Isosurface, 12
- Isosurface statistics, 50
- Jacobi set, 13, 16
- Level set, 12
- Manifold, 9
- Matrix norm, 60
- Morse function, 15
- Morse theory, 14
- Multifield comparison measure, 60
- Multifield comparison measure: Stability, 62
- Multifield Visualization, 3
- Offset operation, 17
- Persistence, 41
- Piecewise-linear functions, 12
- Rectilinear grid, 11
- Reeb graph, 15
- Riemannian manifold, 9
- Riemannian metric, 9
- Scalar fields, Scalar functions, 3
- Sign extended comparison measure, 17
- Silhouette, 30
- Simplex, 11
- Simplicial complex, 11
- Simplification, 16, 19
- Structured grid, 11
- Time varying combustion, 49
- Variation density, 37
- Variation density profile, 42
- Volume rendering, 3