



UNIVERSITY OF KAISERSLAUTERN

Visualization and Analysis of Multifields using Pareto Sets

Author:
Lars S. HUETTENBERGER

Supervisor:
Prof. Christoph GARTH

A thesis approved by the Department of Computer Science of the
Technische Universität Kaiserslautern for the award of the Doctoral Degree

Doctor of Natural Sciences (Dr. rer. nat.)

in the

Scientific Visualization Lab
Department of Computer Science

<i>Date of Defense</i>	December 20, 2018
<i>Reviewers</i>	Prof. Dr. Christoph Garth <i>University of Kaiserslautern</i> Prof. Dr. Gerik Scheuerman <i>University of Leipzig</i> Prof. Dr. Nicolas R. Gauger <i>University of Kaiserslautern</i>
<i>Head of Committee</i>	Prof. Dr. Jens Schmitt <i>University of Kaiserslautern</i>
<i>Dean</i>	Prof. Dr. Stefan Deßloch

*To my family and friends, and all those who are able to follow my thoughts, even at times
when I am too fuzzy to end my current*

University of Kaiserslautern

Abstract

Department of Computer Science

Doctor of Natural Sciences (Dr. rer. nat.)

Visualization and Analysis of Multifields using Pareto Sets

by Lars S. HUETTENBERGER

The focus of this work is to provide and evaluate a novel method for multifield topology-based analysis and visualization. Through this concept, called Pareto sets, one is capable to identify critical regions in a multifield with arbitrary many individual fields. It uses ideas found in graph optimization to find common behavior and areas of divergence between multiple optimization objectives. The connections between the latter areas can be reduced into a graph structure allowing for an abstract visualization of the multifield to support data exploration and understanding.

The research question that is answered in this dissertation is about the general capability and expandability of the Pareto set concept in context of visualization and application. Furthermore, the study of its relations, drawbacks and advantages towards other topological-based approaches. This questions is answered in several steps, including consideration and comparison with related work, a thorough introduction of the Pareto set itself as well as a framework for efficient implementation and an attached discussion regarding limitations of the concept and their implications for run time, suitable data, and possible improvements.

Furthermore, this work considers possible simplification approaches like integrated single-field simplification methods but also using common structures identified through the Pareto set concept to smooth all individual fields at once. These considerations are especially important for real-world scenarios to visualize highly complex data by removing small local structures without destroying information about larger, global trends.

To further emphasize possible improvements and expandability of the Pareto set concept, the thesis studies a variety of different real world applications. For each scenario, this work shows how the definition and visualization of the Pareto set is used and improved for data exploration and analysis based on the scenarios.

In summary, this dissertation provides a complete and sound summary of the Pareto set concept as ground work for future application of multifield data analysis. The possible scenarios include those presented in the application section, but are found in a wide range of research and industrial areas relying on uncertainty analysis, time-varying data, and ensembles of data sets in general.

Acknowledgements

Parts of this thesis are created in cooperation with my colleagues at the University which I like to thank, especially Kathrin and Nils Feige, Christian Heine, and my supervisor Christoph Garth from the University of Kaiserslautern as well as Hamish Carr (University of Leeds), Geric Scheuermann (University of Leipzig) and Michael Böttinger (German Climate Computing Center). A more detailed description of the different cooperation are in the individual thesis chapters.

This work was funded in part by the DFG International Research Training Group 1131 "Visualization of Large and Unstructured Data Sets" and the Marie Curie Actions within the EU FP7 Programm under grant #304099.

Contents

Abstract	v
Acknowledgements	vii
1 Introduction	1
1.1 Motivation	1
1.2 Contribution	2
1.3 Thesis Structure	3
2 Related Work	5
2.1 Topology-based Techniques for Single Fields Data Sets	5
2.2 Visualization of Multifield Data Sets	7
2.3 Topology-based Techniques for Multiple Fields Data Sets	8
2.4 Existing Work on Pareto Optimality	10
3 Mathematical Background	13
3.1 Simplicial Complexes	13
3.2 Single-Field Topology	14
3.2.1 Morse functions	15
3.2.2 Contour Trees and Reeb Graphs	16
3.2.3 Reeb space	17
3.2.4 Morse-Smale Complex	17
3.3 Gaussian functions	19
4 Definition	21
4.1 Motivation	21
4.2 Definition of Pareto Optimality	21
4.3 Global Ascending and Descending Sets	25
4.4 Definition of the Reachability Graph	27
4.5 Implementation for Piecewise Linear Multifields	30
4.5.1 Parallel Computation of the Pareto Set	35
4.5.2 Interactive Computation of the Pareto Set	36
4.6 Discussion	42
4.6.1 Analytical Examples and Visualization Ideas	43
4.6.2 Limitations	46
4.6.3 Runtime	47
5 Simplification	49
5.1 Using Existing Single-Field Approaches	49
5.2 Simplification based on Reachability Graphs (RGS)	50
5.2.1 Interpretation	52
5.2.2 Operation Sequence	54
5.2.3 Implementation	56

5.3	Simplification Based on the Comparison Measure Contour Tree (CTS) .	56
5.3.1	Comparison Measure	57
5.3.2	Contour Tree	58
5.4	Comparison and Discussion of the Simplification Approach	59
5.4.1	Analytical Examples	59
5.4.2	Can Data	61
5.4.3	Discussion	63
6	Comparison	67
6.1	General Comparison with Common Methods	67
6.2	Comparison of the Pareto set with the Jacobi set	69
6.2.1	Alternative Definition of Pareto Sets based on Linear Programs	69
6.2.2	Definition of Jacobi Sets based on Linear Programs	71
6.2.3	Mathematical Background of Jacobi set	71
6.2.4	Alternative Definition of Jacobi Sets based on Linear Programs	72
6.2.5	Translation between Jacobi and Pareto Sets	74
6.2.6	Discussion	75
6.3	Comparison of the Pareto Set with the Joint Contour Net	78
6.3.1	Related Work for Joint Contour Nets	79
6.3.2	Definition of the Joint Contour Net	79
6.3.3	Translation into Directed Joint Contour Net	80
6.3.4	Recognition of Pareto maxima and minima through a dJCN . .	84
6.3.5	Recognition of Pareto optima through a dJCN	85
6.3.6	Definition of ϵ -Pareto extrema	85
6.3.7	Proof-of-Concept Examples over constant 2D-grids	88
6.3.8	Discussion	91
7	Applications	93
7.1	Vortex Detection	93
7.1.1	Scenario 1: Vortex detection in fluid simulations	93
7.1.2	Scenario 2: Atmospheric Hurricane Data	97
7.1.3	Results	103
7.2	Quality Control in Car Manufacturing	103
7.2.1	Scenario Overview: Data Set and Application Goal	104
7.2.2	Definition of the Weighted Pareto Sets	105
7.2.3	Results	106
7.3	Comparing Ensemble Members of Climate Projections	109
7.3.1	Scenario Overview: Application Goal	110
7.3.2	Enhanced Visualization of the Pareto Set	111
7.3.3	Glyph-Based Visualization of Field Contribution	112
7.3.4	Scenario Overview: Global and Regional Climate Ensembles . .	115
7.3.5	Results	117
8	Conclusion	129
8.1	Discussion, Limitation and Relevance	130
8.2	Outlook	131
	Bibliography	133
A	Supplemental materials for Section 7.2	147

B	Curriculum Vitae	151
B.1	Personal Data	151
B.2	Education	151
B.3	Work Experience	152
B.4	Publication List	153

List of Figures

3.1	A set of 2D functions each with a critical points at $x = (0, 0)$. The labels indicate the used equation for $f(x, y)$. (a) shows a maximum, (b) a saddle, and (c) a Monkey saddle. The images are received from the Wikipedia repository [182, 184, 181]	15
3.2	Reeb graph of the height function on a torus. The images are received from the Wikipedia repository [178].	16
3.3	Two height maps and, between them, a contour tree corresponding to both maps.	17
3.4	A illustration from Carr and Duke [29] with two 3D fields f_1 and f_2 . The field $f_1(x, y, z) = z$ is the height map, and the field $f_2(x, y, z) = \sqrt{x^2 + y^2 + z^2}$ is a distance map. The fields are further illustrated through sequences of contour trees of f_1 restricted to an isosurface for f_2 , i.e. $f_2^{-1}(h)$ for some value h , and vice versa.	18
3.5	Example of the Morse-Smale complex for a simple 1D field. The Morse-Smale cells are indicated by the displaced red sections underneath the x-axis.	18
3.6	Gaussian curve $z = e^{-x^2-y^2}$ over a simplicial 2-complex from the free media repository at Wikipedia [179].	19
4.1	Two fields g, h with individual critical points (red circles). The location of Pareto maxima and Pareto minima are indicated by green and red dotted lines, respectively. Pareto optima create two connected areas, colored yellow.	23
4.2	3×3 data points with two sets of values. Pareto maxima, Pareto minima, and Pareto optima are colored green, red, and yellow, respectively. A neighborhood around the point with values $(0, 0)$ is indicated by a hatched area. Ascending paths from this points are indicated by a black line.	26
4.3	An example configuration of Pareto extrema illustrating the reachability graph, the global ascending and descending sets of the connected component $R1$ (shaded red and green), as well as the descending sets of $R2$ (hatched).	28
4.4	A simple example with two sets of scalar values over four vertices and four edges. Vertex v_1 is a Pareto minimum, colored in green, v_3 is Pareto maximal, colored in red, and vertex v_4 and adjacent edges, colored in yellow, are Pareto optimal. v_1, v_3 , and v_4 and all points along the mentioned edges build on connected component S , i.e. one element in $\mathbb{E}(f)$. Note that this component reaches itself via the ascending path along the edges (v_1, v_2) and (v_2, v_3) which results in a loop.	29
4.5	Different configurations with two fields over a selection of triangles to illustrate the concept of ascending (red) and descending sets (green).	31

4.6	Successive clipping of a triangle by half-planes originating from three gradients at the barycenter to determine $H_\sigma^+(x)$ (green region).	32
4.7	The 8 classes of possible triangle intersections with their corresponding 3-bit descriptor. The thick black line marks the intersection, the gray area marks the remaining triangle part and the thin black line indicates possible subtriangles.	33
4.8	Example of a triangle with $f = (f_1, f_2, f_3)$	37
4.9	The Pareto set, colored in strong green, red and yellow according to their Pareto type. In each triangle the ascending and descending set for its barycenter are colored in less intensive colors according to the color scheme in previous Figures.	39
4.9	Pareto set and ascending and descending sets (cont.)	40
4.10	Analytical examples. (a) Pareto extrema of four 2D fields, each with two positive and two negative peaks, and line-integral convolution depicting approximated gradient directions in non-Pareto regions [27]. (b) Pareto extrema along with the global ascending and descending sets for the five Pareto optimal regions. For illustration, some global ascending/descending sets intersections have been annotated with a bit string indicating the relation to the Pareto extremal regions.	44
4.11	Analytical examples. (a) Pareto extrema of eight 3D fields, each based on 3D Gaussian bells with three positive and three negative peaks. The reachability graph is placed in the domain. The nodes are placed at the gravity center of their corresponding elements of $\mathbb{E}(f)$. (b) A graph layout that shows the reachability graph of the multifield in (c). Generated by C.Heine with graphviz's fdp [63].	45
5.1	Illustration of two single-field simplification methods. (a) shows the unsimplified function and indicates the regions that are to be flooded and carved as blue and red areas, respectively. (b) shows the simplified function, if the flooding concept is applied, while (c) depicts the simplified functions after the peak is carved away.	51
5.2	Illustration of the functional change if an edge corresponding to a connection between $R1$ and $R2$ is merged. (a) and (b) show the result of the changes to the ascending and descending sets of $R1$ while (c) provides implementational ideas through a 1D example.	53
5.3	A simple example with two sets of scalar values over four vertices and four edges. Vertex v_1 is a Pareto minimum, colored green, v_3 is Pareto maximal, colored red, and the yellow colored vertex v_4 and adjacent edges are Pareto optimal. v_1, v_3 , and v_4 and all points along the mentioned edges build on connected component S , i.e. one element in $\mathbb{E}(f)$	54
5.4	Illustration of the merge and purge graph operations similar to the work by Suthambhara and Natarajan [121].	58
5.5	An illustration of the reachability graph over a synthetic data set with four scalar fields. (a) depicts the contours of each of the four scalar component functions. The reachability graph, given in (b).	59

5.6	A comparison of the two Pareto set simplification techniques over synthetic data with 1.5% random noise. The top images (b) and (c) show reachability graph simplifications (RGS). The increasing thresholds t are indicated by the image labels. The bottom row (images (e) and (f)) illustrates the comparison measure contour tree simplification (CTS) for two similar thresholds. The first images (a) and (d) present the unsimplified data for the methods.	60
5.7	An illustration of the Pareto set and its reachability graph as an overlay over a real-world data set with four scalar fields, called the can data set. Note that this visualization removes transitively given edges in the reachability graph due to confusion and clutter. See Section 7.1 for more details.	62
5.8	A comparison of the two Pareto set simplification techniques over synthetic data with 0.04% random noise. The top images (b) and (c) shows reachability-graph simplification (RGS) for two increasing thresholds. The bottom row illustrates the comparison measure contour tree simplification (CTS) for two increasing thresholds (images (e) and (f)). The unsimplified data is presented in the images (a) and (d), respectively.	63
6.1	Pareto sets and Jacobi sets for a two-dimensional example with two functions. In both images the same fields are used but in the second image, one of the individual fields is inverted.	76
6.2	Pareto sets and Jacobi sets for a two-dimensional example with three similar Gaussian functions. Subfigure 6.2(a) illustrates one of the three functions as a hight map. The remaining images each show the Jacobi set (white lines) for one of the three possible pairings of the Gaussian functions. All images also show the Pareto set for the multifield containing all three functions together.	77
6.3	The left image shows two functions f and g colored in red and blue, respectively, with an image interval from zero to four. The right image overlays the functions with their discretized versions $round(f)$ and $round(g)$ colored similarly with $round(x) = \lfloor x \rfloor$, the <i>floor</i> function. The yellow colored regions A , B , and C indicate the slabs for the joint level set for $c = (1, 2)$	80
6.4	dJCN and Pareto set for a synthetic example based on a height and a distance function.	82
6.5	dJCNs and Pareto sets for two synthetic examples based on a set of Gaussian functions in 2D and 3D, respectively. Critical slabs in (a) and (b) and the Pareto extrema in (b) and (d) are colored similarly, with minima shown in red for both concepts, and maxima either in blue (dJCN) or green (Pareto sets). Also, (b) shows isosurfaces in blue and gray for the 2D functions.	83
6.6	dJCN and Pareto set for the multifield function based on the U and V component of the LAMPS data set. The grid has over 21000 vertices which results in 95200 tetrahedrons. For the dJCN, the slab width is set to 5 for each underlying function. The calculation of 895 nodes and 1667 edges took around 25 seconds.	84
6.7	An example with Pareto optimal points in slabs which are not marked as such. Pareto maximal slabs are marked in green. Pareto minimal slabs are shown in red, while white areas correspond to regular slabs.	86

6.8	Illustration of the number of slab changes between point x and y for two fields f_1, f_2	88
6.9	Pareto extremal points and slabs colored in green, red, and yellow depending on their type. In (b) the three individual fields are indicated by a selection of contour lines.	89
6.10	Pareto extremal points and slabs colored as in Figure 6.9 but the data set contains an increased number of fields and each individual field has more single-field extrema.	90
6.11	Two 1D fields (top), a dJCN (center) for a 2D extension of the fields, and the corresponding reachability graph (bottom).	90
6.12	Pareto extremal points and slabs colored in green, red, and yellow depending on their type.	92
7.1	Color maps of the scalar fields going from blue for low values to red for high scalar values.	94
7.2	The Pareto set for the fields in Figure 7.1. (a) shows also the corresponding reachability graph, while (b) shows a visualization of all ascending and descending sets of all connected components contained in the Pareto set.	95
7.3	Pareto extrema in a can flow simulation. While (a) depicts Pareto extrema in the 3D volume, the other images focus on an axial slice of the dataset. The different visualizations highlight different aspects. See main text for details.	96
7.4	The path of hurricane Isabel during September 06-19 2003. The black box indicates the spatial area and the white box indicates the time window contained in the data set. The background image containing the hurricane track and the development of its intensity is from a paper by Gautam et al. [66].	97
7.5	Features of hurricane Isabel. The image is from the International Space Station (15. September 2003) [180]).	98
7.6	Pareto set, colored in green, red, and yellow depending in the Pareto extremal type, for air pressure simulated 20 meter above sea level. The multifield contains 21 fields, each representing one hour time steps (see Figure 7.8 and 7.7.	99
7.7	Pareto set as shown in Figure 7.6 with a grayscale map based on the air pressure 20m above sea level at different time steps t indicated by the image labels.	100
7.8	Pareto set as shown in Figure 7.6 with a grayscale map based on the air pressure 20m above sea level at time step 10.	101
7.9	The Pareto set for different numbers of hight layers at time step 20, i.e. around September 16, 20:00 UTC.	101
7.10	The Pareto set for different numbers of hight layers at time step 45, i.e. around September 17, 21:00 UTC.	102
7.11	Selection of distance maps based on 3D scans of the same car trunk lid, displayed color-coded on the quota grid. The colorscale ranges from black (zero deviation from the quota) over red to yellow (high deviation from the quota).	104
7.12	A simple 1D example with three functions f_1, f_2 and f_3 and the corresponding Pareto extrema colored in red for Pareto minima, green for Pareto maxima, and (transparent) yellow for Pareto optima.	106

7.13	The weighted Pareto set on the trunk lid for the six distance maps shown in Figure 7.11 after a Gaussian smoothing with parameters I: 10, FWHM: 6, and ND: 5. Regions are colored based on their weights, going from red (0) over blue (3) to green (6).	107
7.14	Image analogous to Figure 7.13 with different smoothing parameters (I: 20, FWHM: 3, and ND: 5). Note how the oversimplification removed most of the features visible in Figure 7.13.	107
7.15	Image similar to Figure 7.13 (smoothing parameters I: 10, FWHM: 3, ND: 3). Note how most of the Pareto extrema can be found on slopes of the trunk lid.	108
7.16	An illustration of the opening angle visualization based on two functions. Both functions (distance functions based on two different centers) are indicated by contour lines. The Pareto set based in these two function is colored in green while the opening angles are visualized as a grayscale map.	112
7.17	Three glyph designs that could be used to visualize the directions in a vector set. However, none of these designs can highlight outliers sufficiently well.	113
7.18	Modified illustration from [80] of a glyph for three gradients.	114
7.19	Each image shows ΔT_{as} for a different climate projection, as indicated by the image label. While the overlay of a world map provides spatial context, comparison between different projections is obviously difficult using only such a simple multi-window view.	116
7.20	The ISD-map for four global climate projections based on the MPI-ESM model over a 192x96 grid.	117
7.21	A world map with the Pareto set visualization as overlay. The opening angles in each triangle are presented as a transparent grayscale map, with dark gray indicating a small opening angles, and thus high divergence between the projections. Note for example, that incomparability mostly appears over the oceans.	118
7.22	A merge between Figure 7.20 and the smoothed Image 7.21(b). The background contains the map of $\Sigma \Delta T_{as}$ from Figure 7.20, while the Pareto-set visualization is presented as an overlay.	119
7.23	The Kuroshio current is the west side of the clockwise North Pacific ocean gyre. The image is from the free media repository from the Wikipedia website. [183]	120
7.24	The first row, images (a)-(c), show the ISD-map for different subsets of projections based on the same climate model indicated by their subtitles but with different initial conditions. Images (d)-(f) show a world map and the Pareto set visualization for the same subsets.	121
7.25	Image (a) shows a merge between the $\Sigma \Delta T_{as}$ -values (Figure 7.24(c)) and the Pareto sets and opening angle visualization (Figure 7.24(f)) for the climate projections based on CSIRO-MK3-6 over the United States. In (b), two larger Pareto extremal regions are selected and highlighted in a slightly darker orange, to identify possible projections, responsible for this large topological divergence. Note that the $\Sigma \Delta T_{as}$ -color map is removed in (b) to make it easier to connect the glyph segments to their corresponding projection.	122

7.26	A section of the United States with the Pareto set and opening angle visualization for the CSIRO-MK3-6.i projections, $1 \leq i \leq 8$. In (a), a selection of eight Pareto extremal regions is chosen to analyze, which projections are responsible for the topological divergence around this section of the USA. Image (b) shows the same section after CSIRO-MK3-6-0.1 is removed from the set of projections. Note the decreasing extent of the Pareto-optimal regions.	122
7.27	A detailed visualization of a region, selected in Figure 7.26(a). The numbers i in each glyph segment indicate the corresponding projection CSIRO-MK3-6-0.i.	123
7.28	Grayscale map of T_{as} for April 2020 as projected by six modeling systems employed in CMIP3. References to the models can be found in [22]. Darker greys refer to lower and lighter greys to higher temperatures.	124
7.29	The Pareto set for 40 projections. The box indicates a region where further investigation is done, see Figure 7.30.	124
7.30	Using glyphs to explore diverging model behavior.	125

List of Tables

7.1	Hurricane Isabel data produced by the Weather Research and Forecast (WRF) model, courtesy of NCAR and the U.S. National Science Foundation (NSF) [106].	98
7.2	Data set characteristics and run time in seconds for different Pareto set calculations	126
A.1	Parameter settings for the Gaussian smoothing used to create different weighted Pareto set visualizations are indicated by the image ids in the first column.	147

List of Pseudocode Segments

4.1	Recursive pseudocode for the ascending or descending set of a point in a triangle.	33
4.2	Pseudocode to identify the Pareto set in a simplicial complex.	35
4.3	Iterative pseudocode for the ascending or descending set of a point in a triangle.	41

List of Abbreviations

\mathbb{M}	A d -dimensional manifold	13
\mathcal{S}	A d -simplicial complex	13
iff	if and only if (\Leftrightarrow)	19
$x \succeq y$	x (weakly) dominates y	22
$x \succ y$	x strictly dominates y	22
$x \not\succ y$	x and y are incomparable	22
$U(x)$	An open neighborhood containing x	22
$\mathbb{P}^+(f)$	Set of Pareto maxima for the multifield f	23
$\mathbb{P}^-(f)$	Set of Pareto minima for the multifield f	23
$\mathbb{P}^o(f)$	Set of Pareto optima for the multifield f	23
$\mathbb{P}(f)$	Pareto set for the multifield f	23
$H_\sigma^+(x)$	Local ascending set for point x restricted to σ	24
$H_\sigma^-(x)$	Local descending set for point x restricted to σ	24
$H_\sigma^\pm(x)$	Set of all comparable points to x restricted to σ	24
$x \rightsquigarrow y$	x is connected to y by an ascending path	25
$\mathbb{E}(f)$	Connected components in $\mathbb{P}(f)$	25
p^+	Ascending path	25
p^-	Descending path	25
$\mathcal{C}^+(x)$	Global ascending set for point x	25
$\mathcal{C}^-(x)$	Global descending set for point x	26
$RG(f)$	Reachability graph of the multifield f	27
RGS	Reachability graph-based simplification.....	50
$\alpha^+(x)$	Opening angle of the ascending cone for point x	55
$\alpha^-(x)$	Opening angle of the descending cone for point x	55
$RG_w(f)$	Weighted reachability graph of the multifield f	55
CTS	Contour tree-based simplification.....	56
\mathcal{P}_p^+	Linear program corresponding to $p \in \mathbb{P}^+$	70
\mathbb{J}	Jacobi set f	71
\mathbb{J}^+	positive Jacobi set f	71
\mathbb{J}^-	negative Jacobi set f	71

\mathcal{D}_p^+	Linear program corresponding to $p \in \mathbb{J}^+$	73
JCN	Joint Contour Net.....	79
dJCN	directed Joint Contour Net.....	81
$CDist_f$	Chebyshev distance	86
$\mathbb{P}_\epsilon(f)$	ϵ -Pareto set	86
$\mathbb{P}_w(f)$	Weighted Pareto set	106
ISD	Inter-modul standard deviation [168]	109
β^+	Quantized version of α^+	111
CMIP	Coupled Model Intercomparison Project	116

Chapter 1

Introduction

1.1 Motivation

Based on technological advancement in computational and sensory capacity over the past years, the data volume in scientific research and applications has increased steadily and will most likely follow in this trend. This increase affects not only the number of data sets but also their complexity in terms of size and resolution, thus making effective analysis and visualization more and more difficult. The increased complexity also leads to a transition from single- to multifield data sets. Instead of a single field, a set of fields is given to provide, for example, different domain resolutions of the same data, different variables or time-steps of the same simulation, or, to compensate uncertainty, the results of repeated simulation or measurement runs [81]. Analysis and visualization tools aim to compare or combine the individual fields to support domain experts as well as regular users in tasks like identification and exploration of common behavior and structures, outliers and trends in the data.

To guide these tasks when the amount or complexity of the data is too high to simply presenting it in one image per individual field, a general approach is to provide combined, comparing or abstracted views on the data [131]. Following this goal for single- and multifield data alike, topology-based analysis and visualization methods can be utilized to describe and organize structures in the data, to highlight data subsets, and to guide data exploration. Those approaches are especially applicable in scenarios for which some level of relation between the individual fields can be assumed as for example the case for multifield or ensemble data sets.

Applications for such approaches can be found in a wide area of scientific research and real world scenarios. Examples, some which are included in the application section of this work (Chapter 7), can be found in:

Fluid simulation Abstracted flow visualization helps to provide an overview and to understand complex, time-varying fluid behavior in large, complicated vector fields.

Vortex detection In multi-criteria analysis for vortex detection in combustion chambers, information from different vortex criteria are combined and structured to find common ground and to compensate for uncertainty in the individual criteria.

Quality control Automated, sensory measurements create large quantities of data sets to ensure the quality and durability of machines and products. However, sensor errors are always possible such that it is necessary for analysis purposes to distinguish between common, global trends in the data and local errors.

Environmental studies This includes weather forecast, analysis of Hurricane structures, or climate change studies. In general, different projections of prediction models or different time steps are compared to identify local and global trends and outliers, and thus to improve the existing models further.

Note that application domains involve scalar-, vector-, or tensor fields, combinations of those, as well as fields with different resolutions. This thesis, however, focuses only on scalar fields over the same domain and resolution.

In the context of multifield scalar fields, topology-based approaches include concepts like the Reeb space, Jacobi set, and Joint Contour Nets. Those are well studied but have different limitations. The number of individual fields is restricted by the number of domain dimensions, the created structures are high-dimensional and thus hard to visualize in 2D or 3D or are invariant to the orientation of the individual fields.

The study of novel multifield approaches is therefore still an open task, especially since the assumption of common behavior in the data does not hold in general. In such cases, unrelated features in the individual fields result in separate, local and small-scale structures in the multifield, drastically increasing its complexity and making abstraction and even simplification of the multifield necessary.

1.2 Contribution

In this dissertation, the design, visualization and evaluation of an approach for continuous scalar multifield data, the *Pareto set*, is presented. This concept avoids limitations of existing concepts and yields a topology-based structures such that an abstract visualization and exploration of multifield data becomes possible.

To provide a complete introduction of Pareto sets and its related structures, the following contributions are reflected on in the individual chapters of this dissertation.

- The main contribution is the introduction and definition of the Pareto set concept itself which is used as a framework throughout the dissertation.
- Additionally, guidelines for an efficient implementation to identify the Pareto sets in piecewise linear multifields in $2D$ and $3D$ domains like simplicial complexes as well as visualizing strategies for them are provided.
- To discuss options to simplify and organize the Pareto set for easier visualization and to understand internal structures and connectivities within the Pareto set, a related structure, the *reachability graph* is introduced and defined.
- To evaluate the Pareto set concept, this thesis provides thorough comparison with existing topology-based multifield concepts. The results are utilized to further improve the Pareto set implementation.
- The usability of the Pareto set is tested using several real-world application scenarios, as mentioned above.
- In context of these applications, further extensions and improvements of the Pareto set visualization are provided.

In summary, this dissertation provides all necessary parts to understand and calculate the Pareto set as well as its related structures and visualizations in the context of scalar multifields. Application examples present to the capabilities of the concept, show and discuss when and how it can be applied, and how to interpret and improve the results. Thus, the thesis provides all parts to place the Pareto set concept firmly into the research field of topology-based multifield analysis, visualization, and applications.

1.3 Thesis Structure

The Pareto set concept introduction is separated in three main chapters: theoretical definitions and considerations, comparison with existing concepts, and application examples.

To follow the theoretical definitions, related scientific works and mathematical background for the Pareto set concept, respectively are introduced in the first two segments of this thesis (Chapters 2 and 3) excluding the introduction.

Chapter 4 contributes the definition of the Pareto set as well as its inner structure, the reachability graph, based on an existing concept of Pareto optimality from multi-criteria graph optimization. The definition is applicable to continuous scalar fields over the same domain with an arbitrary number of individual fields and domain dimensions.

With respect to applications, the thesis focuses on piecewise linear scalar fields and provides instructions for multi-threaded implementations in 2D and 3D domains. For these cases, equivalent definitions of the Pareto set that further reduce the computational effort are introduced as well as a framework to calculate the Pareto set at interactive speeds, allowing for inclusion and exclusion of individual fields without a complete recalculation. For each improvement, the resulting implications and limitations, for example with respect to data sets, running time, visualization are discussed. Hence, in Chapter 4 Pareto set and the reachability graph are defined to help users to understand and, with the provided guidelines, to implement this concept for their own applications.

In the next segment (Chapter 5) the high complexity of multifield data and the susceptibility of multifield analysis and visualization concepts, including Pareto sets, towards small-scale structures, loosely denoted as "noise" is considered. Existing approaches are evaluated and a novel simplification method for Pareto sets is introduced based on the reachability graph.

The second part of the introduction is contained in Chapter 6. In this comparison chapter, the relation between the Pareto set and other topology-based multifield concepts is studied. Besides a general discussion about advantages and disadvantages of the Pareto set towards other commonly used methods and related work, the work focuses on the relations with Jacobi sets and Joint Contour nets. Both concepts are well studied and provide efficient implementations as well as relations to other multifield methods. This allows for a transitive relation of the Pareto set to these other topological structures, including the important Reeb space.

This places Pareto sets firmly into the context of multifield approaches, helps users to decide when Pareto sets are suitable or even better than other concepts for their application, and highlights improvements through the combination of different methods.

In Chapter 7 application examples of the Pareto set in real-world scenarios are addressed. The potential of Pareto sets as a visualization tool is shown in a variety of different problems and, in each scenario, it is discussed how the original Pareto set definition and its visualization is used and improved for data exploration. In addition to these discussions, throughout the chapter arising open tasks for further improvement are summarized.

Improvements for the Pareto set concept that are included in this thesis are glyph-based visualization, weighted Pareto sets based on single-field topological structures and time-varying information, and domain decomposition based on the robustness of non-Pareto points.

With this chapter, the wide range of applications of the Pareto set visualization is emphasized, which helps users to consider possible scenarios and improvements of that concept.

The last segment (Chapter 8) contains the conclusion of the dissertation with a summary of the limitations and possibilities of the Pareto set concept and an outlook on future work.

Chapter 2

Related Work

Throughout this dissertation, a novel topology-based approach for analysis and visualization of multifield scalar fields is presented with focus on two- and three-dimensional domains. To better follow the overview of related work that places Pareto sets in a proper context, the background section is structured in non-topology-based multifield, topology-based single-field, and topology-based multifield techniques, i.e. methods that use concepts and ideas from computational topology [51] to visualize or enhance the visualization of scientific data.

2.1 Topology-based Techniques for Single Fields Data Sets

The extraction of topological structures not only provides a meaningful summary of the data, but these structures are also often used to drive exploration of the data in interactive systems, or inform parameter choices in later visualization phases. In this context, multifield approaches are usually based on well-studied methods for single fields.

The basis for many approaches are critical points, like minima and maxima in scalar fields [157, 139] that can be extended by additional concepts to show the connection between those points, the so-called topological structure. Contour tree and Reeb graph [48] contract connected components of isosurfaces to points in a graph structure and translate the adjacency of isosurfaces to the adjacency of those points. Additional information, like level set volume or boundary size can be added to the graph as edge weights to transport additional information [102]. Ushizima et al. [166], for example, use a Reeb graph enhanced with thickness information to detect and visualize material flow and pockets in rock formations to study ways to capture CO₂ and reduce the carbon dioxide concentration in the atmosphere.

Carr [31] presents the flexible isosurface interface where individual contours in a 3D view can be manipulated by moving single points in a drawing of the data's contour tree. As an application, this interface is used to mark and color different contours to identify materials and thus different anatomical parts from MRI brain scans in medical research [33, 98]. Weber [173] extends this approach to volume rendering by assigning individual contour tree edges their own transfer function. Thomas and Natarajan [160] identify symmetric structures in the extremum graph to identify symmetric structures in the corresponding scalar fields, for example to highlight patterns in the data.

Bremer et al. [23] use Jacobi sets and Weber et al. [172] use Reeb graphs to track volumetric features across time. Bremer et al. [24] track burning flames in a combustion simulation on the basis of different physical quantities. The tracked features,

such as critical points, can be used to efficiently update contour trees over time, in time-dependent data fields [113].

Similar concepts can be applied to other data types. For vector fields, Morse-Smale complexes [74, 149] are an example. In this context a partition of the domain into regions is constructed, clustering integral lines with the same origin and the same destination. However, all approaches consider only one single field. Gyulassy et al. [73] combine the domain partition with user defined borders to assist physicians in histology studies, which is one possible application. This technique requires, in the context of the application, useful input to produce meaningful results.

Gyulassy et al. [78] detect core structures (skeletons) in porous materials through a topologically simplified Morse-Smale complex of a signed distance field. This supports material scientists to perform quality and quantity analysis of materials for pores, tunnels, and holes.

Garth and Tricoche [64] present feature- and topology-based techniques, like cutting-plane topology and spiral-critical points that are tracked through a flow field. Corresponding visualizations are used, for example, to show general motion patterns inside combustion engines.

Besides providing abstract data visualization and supporting interactive data exploration, such graph structures can be used for data simplification. In many applications, structural richness may result from noise or sampling inaccuracies, and robust topological visualization methods should ideally be impervious to such phenomena. Hence, for all those data types, simplification techniques have been proposed and studied, with the aim of allowing topological visualization of structurally rich data. A common approach to guide topological simplification is persistent homology [49], for example of Reeb graphs [162, 50, 128] or for Morse-Smale simplification [72, 41, 138].

In data mining, the simplification of merge trees based on density functions can be an alternative to clustering [126]. Here, Oesterling et al. generate a density field from a set of point data, that are otherwise visualized through methods like scatter plots, hyperboxes, or parallel coordinates [10, 3, 94], since they lack a field, placing the point values in a continuous domain. Surveys, regarding the visualization of multivariate point data are, for example, found in work by Chan or Xie et al. [35, 189].

The concept of *persistence* in general captures the existence interval of topological features as scalar values change. For two adjacent nodes in a Reeb graph, this corresponds to scalar value difference between the corresponding critical points [50]. Under the assumption that unimportant features or those resulting from noise have low persistence, i.e. tend to disappear under slight changes to the data, they can be safely omitted from visualization. Such simplification must be consistent in the sense that it captures the topological structure of hypothetically modified data that is close to the original. Recent work introduces a flooding-based method that “fills” a feature up without introducing new ones [162].

2.2 Visualization of Multifield Data Sets

A large drawback of single-field data is that it cannot address uncertainty in simulation processes or small-scale errors in the measurement devices. Such obstacles are found in a series of applications, like hurricane structure or paths analysis, volcano eruptions in atmospheric research, and fluid dynamics [67, 105, 9]. Approaches like presented by Gyulassy et al. [73] integrate user input into the visualization that can be used to address uncertainty through expert domain knowledge. However, user input is not always available or suitable in the context of the application. Hence, multifield data are used to compensate for that.

To visualize multifield data, several techniques exist, although not all visualization methods use topological structures. A general alternative is to juxtapose modalities in multiple linked views [133, 185, 186], presenting each individual field in an individual window, for example to study the flow behavior in different engine parts [25]. Each image can furthermore be enhanced with topological structures, like the extremum graph [105]. However, in these approaches, interactions between modalities are not explicitly expressed in the images but must be explored by the user. This is a difficult task, especially if the number of fields is high.

Two possible approaches to support the exploration are either to integrate all fields into a single volume or to provide comparison measurements between the fields to allow for field traversal and clustering. Kniss et al. [103] propose to use multidimensional transfer functions for rendering volumes based on multiple input data like multi-modal MRI scans [167]. For multifield data, to reduce the complexity of the transfer functions, Kniss et al. suggest to use Gaussian transfer functions instead.

Urness and Interrante [165] discuss the visualization of vector fields by presenting their stream lines in a single image and provide guidelines for streamline placement and overlapping. Based on this work, Urness et al. [164] compile strategies to overlay streamline features visualization and glyph-based flow visualization to create a single images from multiple vector fields. Their presented examples and discussions, however, suggest that this concept only works for two input fields.

Woodring and Shen [188] use boolean set operations to integrate several volumes together into one volume for visualization. Sauber et al. [143] show global correlations of multifields where the user can select a subset of fields to find regions that behave coherently. Botchen et al. [18] integrate extracted features from multifields using fuzzy logic to guide isosurface selection in a single view. Nagaraj et al. [122] define a gradient-based similarity measure that is visualized to show field coherence. In subsequent work, Nagaraj and Natarajan [120] propose a new measure – the so-called variation density function – to guide isosurface selection.

Akiba and Ma [2] suggest to couple information visualization with volume rendering to visualize time-dependent, multifield scalar fields. In another combination of information and scientific visualization techniques, Jänicke et al. [96] use brushing and linking of a graph structure called attribute cloud to visualize multifield data. Lampe et al. [107] transform volumes in a neighborhood-preserving fashion to curves that can be laid out and analyzed in parallel. Fuchs et al. [62] use an evolutionary search algorithm to assist hypotheses formalization in multifield, volumetric

data. Rübel et al. [140] consider multivariate data as point data and use parallel coordinates to visualize and select groups of points with similar attributes.

Unsteady multifields are an extremal case of uncertainty, where the data is not given consistently through a function but as a point set. As with univariate point data, it is possible to use density driven clustering and decomposition to create combined visualizations [95]. Another idea to create continuous maps that contain the combined information of ensemble data is to consider an approach by Baruque et al. [8] that uses ensemble methods with topology preserving models to train a continuous maps.

Liu et al. [112] use longest common subsequence approach to calculate a distance between pathlines of vector field ensembles. These distances are used to guide clustering or calculate similarity field to identify regions with low or high uncertainty in the model that created the ensemble. Cox et al. [42] compare explicit and implicit visualization of an ensemble of hurricane path predictions in a user study. They conclude that an implicit representation, here an error cone, can lead to misinterpretation of the uncertainty of the data.

To visualize high-dimensional data, Weber et al. [174] creates an ensemble of low-dimensional landscapes all with the same contour tree as the high-dimensional data. Harvey and Wang [82] define distances between these landscapes and use nonlinear dimensionality reduction to project the ensemble member onto a 2D plane and allow their exploration.

For further reading, Bürger and Hauser [26] give an overview of further multifield and multifield visualization techniques.

2.3 Topology-based Techniques for Multiple Fields Data Sets

Multifield visualization techniques as presented above have different drawbacks. Methods like overlays or linked views cannot handle large numbers of individual fields. Therefore, approaches using transfer functions, correlation or average maps, etc., summarize multiple fields into a single image. However, large, high-dimensional, or complex data still need some kind of abstract view or structuring to allow for efficient data analysis and exploration. As with single-field data, topology-based approaches exist and aim to provide abstracted views onto the structures inherent in multifield data. Those approaches are often extensions of single-field methods into more complex concepts instead of single-field methods applied to summaries, like the correlation or average map. The latter obviously loses information and transparency in the intermediate, compacting step.

In this context, an important extension towards multifield field visualization is the Reeb space [53, 141]. Analogous to the Reeb graph, fibers, isosurfaces based on all variable values are contracted to points in a high-dimensional construct. Its dimension is linked to the number of individual fields, thus not suited for visualizing a high numbers of fields. Different approaches aim to reduce the dimensional complexity.

Carr and Duke [30] expand the fibers into so-called slabs, which are connected regions of points equal in quantized isovalues for all fields and then generate a graph that encodes adjacency of these slabs. This so-called joint contour net represents the multifield topology for the chosen quantization and is presented using graph drawing techniques. Zhao et al. [67] apply this concept to visualize a multivariate data set containing values like precipitation, pressure, and temperature for hurricane analysis. The results are still approximations and require a good choice of the slab size to produce good results.

As an application in the area of uncertain data, ensembles of repeated simulation runs or measurement operations are analyzed collectively to compensate for uncertainty or inaccuracy in the individual data sets, thus creating multifield data. Mihaela and Westermann [115] define a confidence interval based on the Hessian matrix for each domain point to provide a likelihood for critical points. Similar, Günther et al. [71] build a probability distribution function based on the ensemble data to provide a upper and lower value bounds and therefore level set value intervals. Günther et al. use those to create structures similar to split trees with limitations regarding the probability distribution and saddle points for highly complex data sets.

Edelsbrunner and Harer [52] define the Jacobi set of Morse functions as the set of critical points of one function's restrictions to the fibers of the other functions augmented by the critical points of the single functions. That method is only applicable as long as the number of individual fields does not exceed the dimension of the domain manifold. Hence, Jacobi sets can be useful for bivariate functions in 2- or 3D domains, for example to calculate the more complex Reeb space efficiently [161]. Mascarenhas and Snoeyink [113] as well as the implementation later in the thesis indicate that domain triangulation resulting in simplicial grids create spurious elements in the Jacobi set.

As stated before, Edelsbrunner and Harer [52] also suggests to use Jacobi sets to track critical points in time-dependent scalar fields as an application for unifold data sets, considering time as an additional fields thus creating a bivariate/multifield scenario. In this case, the Jacobi set consists of one-dimensional lines, and their length can be interpreted as life span. Bremer et al. [23] use the life span as a base for a simplification approach on Jacobi sets.

Also, Jacobi sets are used to define time-varying Reeb graphs of single scalar fields on a 3-sphere [58] and later on 4- or 5-manifolds [57], and culminate in the definition of Reeb spaces [53]. Edelsbrunner et al. [55] create a similarity measure based on the difference in the field gradients, an extension to the definition of Jacobi sets with the same restrictions.

Schneider et al. [145] in an extension of work by Schneider et al. [146] propose the following approach: for each given scalar field they compute its largest contours, i.e. maximal contours that contain only one critical point, then compute similarities between largest contours of all fields based on a normalized spatial-overlap measure and store this information in a weighted graph to which graph clustering is applied. Largest contours are considered to be volumetric features and graph clusters can therefore be interpreted as sets of features that are consistent across fields. However, the similarity graph and its clustering are no typical topological structures and the

information is not integrated, e.g. to give a subdivision of the domain.

For multiple vector fields, Morse decomposition is proposed by Szymczack [156] in connection with supertransition graphs. Here, a given grid is refined into smaller subgrids, each represented by a node. Any vector inducing movement from one subgrid to another is recorded as an edge in the graph. Again, the simplicity of the graph can be controlled through by refining the subgrids. While this can be applied to almost any problem, a coarser grid might not represent important topological information. A simplification through removal or merge of unimportant information would therefore be desirable.

An approach by Tricoche et al. [163] for single vector fields that also works for multifield data is to identify critical points in each function and then cluster them in the common domain to reduce the structural complexity. The clustering is nonetheless based on geometric distances between the points and not on any topological structure.

However, so far only limited results are discussed toward the simplification of multifield structural visualization. A promising method is the simplification of Jacobi sets through a local difference measure, described by Suthambhara and Nataraajan [121] for two functions f and g on a 2-manifold embedded in \mathbb{R}^3 . Their measure is based on $\nabla f(x) \times \nabla g(x)$ to describe the similarity between the two functions. Although, it might be possible to expand this measure to multiple functions, the limitation for Jacobi sets still exists, namely $k \leq d$.

For further related work, Heine et al. [84] present a detailed survey regarding topology-based methods for single- and multifield data. Kehrer and Hauser [100] provide a survey regarding the general visualization of multifield.

2.4 Existing Work on Pareto Optimality

While the definitions used in this thesis are based on concepts introduced by Stadler and Flamm [152], the concept / theoretical foundation of Pareto optimality was already introduced in earlier work. .

Stadler and Flamm refer in their definitions to Ehrgott and Gandibleux [59] who define Pareto optima in the context of (multiobjective) combinatorial problems with a finite set of feasible configurations. Pareto optimality is calculated globally as an optimum under all configurations. Stadler and Flamm extend this to multiobjective optimization problems over graph structures in which the different configurations are nodes in a graph. Edges restrict traversal between the different configurations and allow the definition of neighbors and paths in the configuration space. Similar to Ehrgott and Gandibleux, Stadler and Flamm define Pareto optimal configurations with respect to multiple objectives, though locally, and also introduce concepts like the *barrier tree*.

With an application focus in economics, Gerard Debreu [44] introduced Pareto optimality in 1959 in the context of multiobjective optimization over an open set in \mathbb{R}^l to model pure exchange economy markets. Smale [150] adapts the definitions to smooth manifolds in a mathematical attempt to generalize Morse theory [116]

from single to multiple fields, i.e. creating a definition of topology for multifields. In his work, Smale defines critical points, Pareto optima, based on the derivative of the multifield at these points and outlines the idea of "Morse inequalities" for multiple fields. Wan [169] follows those outlines, proving the concept of those Morse inequalities and creating a structure similar to Betti numbers [116] for two functions over a compact manifold W^n ($n \geq 2$).

De Melo [43] also bases his work on Smale's introduction of Pareto optimality for manifolds and defines a stratification of the set of Pareto optima. Stratification is defined as a cover of the Pareto set by a finite number of pairwise disjoint connected sub-manifolds that fulfills the regularity conditions [114]. De Melo proves in his work that this stratification is stable under small perturbations of the corresponding multifield with at most $n \leq 3$ fields.

Note that all those introductions of Pareto optimality are from a mathematical background with a focus on improving either optimization problem solving or the theory behind multifield topology. Neither group has visualization or implementation as a major focus. Hence, aspects like real-world application scenarios, simplification methods or a comparison with other visualization approaches are not considered.

Chapter 3

Mathematical Background

To provide a foundation for the definitions in Chapter 4 and 5 but also to understand the techniques in the comparison section (Chapter 6) the current chapter introduces the needed mathematical framework and definitions for multifield data. More complex concepts like the Jacobi set or multi-modal optimization are revisited directly when needed.

In general, multifield data [81] are sets of individual fields over the same domain \mathbb{M} such that the data is defined as $f = (f_1 \dots f_n)$ with n being the number of individual fields and each field $f_i : \mathbb{M}_i \mapsto \mathbb{I}_i$ with $\mathbb{M}_i \subseteq \mathbb{M}$ and some image space \mathbb{I}_i with the only restriction that a shared domain \mathbb{M} exists such that $\mathbb{M}_i \subseteq \mathbb{M}$ holds for all i . With this general definition, several subtypes of multifield data are possible. However, in this work, the focus is on multivariate and ensemble data over the same domain such that $\mathbb{M}_i = \mathbb{M}_j$ for all $1 \leq i, j \leq n$. In multivariate data, related variables are computed or measured, while in ensemble data, each field is a separate output of a single or repeated computational or measurement process. Related variables or for example functions that are based on the same scenario, or their features have a similar semantic. Hence, throughout this thesis, multifield refers to multifield data over the same domain with special focus on multivariate and ensemble fields. Other types, like multiscale, derived or spectral data exist but are not in the focus.

3.1 Simplicial Complexes

While all theoretical definitions in this dissertation apply to manifolds as a domain $\mathbb{M} \subseteq \mathbb{R}^D$, $d \leq D$, for real-world scenarios a restriction to simplicial complexes is necessary to allow machine-based calculation. A d -manifold is a topological space that locally resembles a d -dimensional Euclidean space [177] near each point. More precisely, each point of an d -dimensional manifold has a neighborhood that is homeomorphic to the Euclidean space \mathbb{R}^d . Two-dimensional manifolds are also called surfaces. Examples include the plane, the sphere, and the torus [110, 109].

To efficiently save and process the domain, manifolds are represented using simplicial complexes. A simplex is a generalization of the notion of line, triangle or tetrahedron to arbitrary dimensions. Specifically, a d -simplex is a d -dimensional polytope which is the convex hull $\{a_0x_0 \dots a_dx_d \mid \sum_{i=0}^d a_i = 1 \wedge \forall_{i=0}^d a_i \geq 0\}$ of its $d + 1$, affinely independent vertices $x_i \in \mathbb{R}^d$. A simplicial complex \mathcal{S} is a set of simplices that satisfies the following conditions:

- Any face of a simplex from \mathcal{S} is also in \mathcal{S} .
- The intersection of any two simplices $\sigma_1, \sigma_2 \in \mathcal{S}$ is either \emptyset or a face of both σ_1 and σ_2 .

A simplicial d -complex \mathcal{S}^\square is a simplicial complex where the largest dimension of any simplex in \mathcal{S}^\square equals d . For instance, a simplicial 2-complex must contain at least one triangle, and must not contain any tetrahedra or higher-dimensional simplices. A common example is a triangulated plane.

To efficiently save and process a field over a simplicial complex \mathbb{S}^d , the data values are only stored at the vertices. This is called a simplicial map. The field is then approximated and expanded through bilinear interpolation over the simplicial complex. More precisely, assume a point $x \in \mathbb{R}^d$ lies inside a d' -simplex σ , $d' \leq d$ such that x can be described through barycentric data $x = \sum_{i=0}^{d'} b_i v_i$, with $b_i \geq 0$ for all $0 \leq i \leq d'$ and v_i the vertices of σ . Then, the value at x is defined by $f(x) = \sum_{i=0}^{d'} b_i f(v_i)$, resulting in a piecewise linear field. Note that this approximation is found in many real-world application, including all applications scenarios described in Chapter 7.

For more details on manifold and simplicial complexes see for example work by Munkres [119] or by Hatcher [83], respectively.

3.2 Single-Field Topology

Topology-based methods for multivariate data are often either based or related to topological methods for single-field data. The following two subsections therefore introduce major visualization and analysis concepts for scalar and vector fields, respectively. In both cases, the field separation and continuous deformation is applied to reduce the visual complexity while preserving its structure including properties like neighborhoods and connectedness.

Both concepts use the general definition of critical points, i.e. points with zero gradient ($\frac{\partial}{\partial x} f(x) = 0$) [108]. Using the second derivative, different types of critical points can be distinguished. A point $x \in \mathbb{M}$ is defined a maximum if $\frac{\partial^2}{\partial x^2} f(x) > 0$, a minimum if $\frac{\partial^2}{\partial x^2} f(x) < 0$, and otherwise a saddle if $\frac{\partial^2}{\partial x^2} f(x) = 0$.

Alternatively, a maximum x for the single field $f : \mathbb{M} \mapsto \mathbb{R}$ is defined by the scalar value in its neighborhood. In the case of manifolds, the neighborhood can be defined by distance such that a point is a maximum, iff

$$\exists_{\epsilon > 0} \forall_{y \in \mathbb{M}} |y - x| \leq \epsilon \Rightarrow f(x) \geq f(y)$$

and a minimum iff

$$\exists_{\epsilon > 0} \forall_{y \in \mathbb{M}} |y - x| \leq \epsilon \Rightarrow f(x) \leq f(y).$$

Note that Fermat's interior extremum theorem Theorem citebartle2000introduction shows the equality of both definition of maxima and minima, respectively, as it states that if x is a extremum (minimum or maximum) and the derivative f' at x exists, than $f'(x) = 0$ has to hold. For an illustration of critical points, see Images 3.1(a) for a maximum and Images 3.1(b) and 3.1(c) for two examples of saddle points.

3.2.1 Morse functions

As a general assumption in the following topological methods, it is supposed that each field is based on a Morse function $f : \mathbb{R}^d \mapsto \mathbb{R}$. Those functions fulfill three conditions [51]

1. f is smooth
2. All critical points, have distinct function values
3. All critical points are non-degenerate (i.e. $\det(\text{Hessian}(p)) \neq 0$)

Hessian is the $d \times d$ square matrix H of second-order partial derivatives with elements $H_{i,j} := \frac{\partial^2 f}{\partial x_i \partial x_j}$. As a direct implication of (1) and (2), f does not contain flat areas, i.e. neighborhoods where f is constant. With condition (3), Theorem 5 by D. Gauld [65] is applicable:

Theorem 3.2.1. *Let 0 be a non-degenerate critical point for $f : \mathbb{R} \mapsto \mathbb{R}$. Then there is a diffeomorphism θ that maps a neighborhood of x in \mathbb{R}^d onto another such neighborhood with $\theta(0) = 0$, and there are numbers $c_i = \pm 1$ ($i = \{1, \dots, d\}$) such that for all $z = (z_i)$ in the domain of θ ,*

$$f\theta(z) = \sum_{i=1}^d c_i z_i^2 + f(x).$$

A diffeomorphism θ is invertible functions, such that both θ and θ^{-1} are defined and differential, for example deformations like rotation, translation or scaling. Thus, after some deformations, the neighborhood around each critical point in a Morse function f is parabolic. For 2D, there are four different configurations for the c_i . The images (a) and (b) in Figure 3.1 illustrate two of them. Image (c), however, shows a so-called *Monkey saddle* with a degenerated critical point in $x = (0, 0)$. Note that the neighborhood around x cannot be deformed into a parabolic shape.

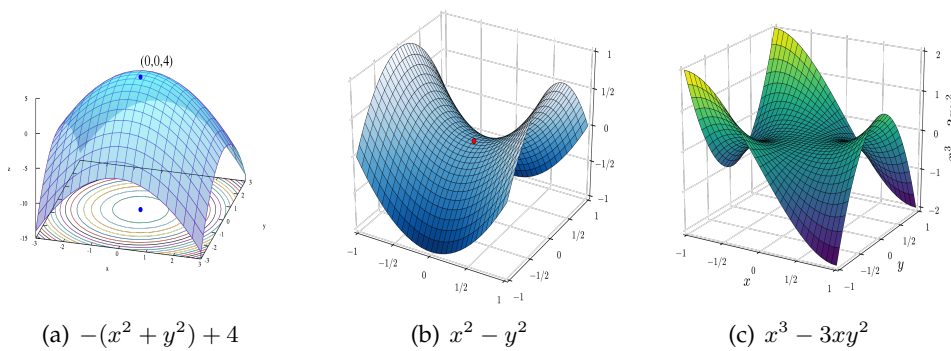


FIGURE 3.1: A set of 2D functions each with a critical points at $x = (0, 0)$. The labels indicate the used equation for $f(x, y)$. (a) shows a maximum, (b) a saddle, and (c) a Monkey saddle. The images are received from the Wikipedia repository [182, 184, 181]

Hence, with a Morse function, for every point x there is a neighborhood such that every direction from x can be described as monotone increasing or decreasing. These implications are used in the implementations and proofs, see for example

Lemma 4.3.1. Resulting limitations are discussed at the end of Chapter 4 after the Pareto set concept is thoroughly introduced. Hence, for the remainder of this thesis, it is assumed that all individual fields in a multifield are Morse functions.

3.2.2 Contour Trees and Reeb Graphs

For a continuous scalar field $f : \mathbb{M} \mapsto \mathbb{R}$, with $\mathbb{M} \subseteq \mathbb{R}^d$ some manifold, the *Reeb graph* and the *contour tree* are based on the connected components of the level sets $f^{-1}(c) = \{x \in \mathbb{M} \mid f(x) = c\}$ for each $c \in f(\mathbb{M})$. Two points $x, y \in \mathbb{M}$ with $f(x) = f(y)$ are in the same component if a function $p : [0, 1] \mapsto \mathbb{M}$ exists such that

- p is continuous,
- $p[0] = x$ and $p[1] = y$, and
- $\forall_{z \in [0,1]} f(p[z]) = f(x)$, i.e. p lies in the level set.

As the *isovalue* c changes the level set $f^{-1}(c)$ can evolve, in other words its components can split, merge, appear or vanish. A graph-based representation, the Reeb graph, illustrates how the level set changes with the isovalue. Leaf nodes represent the creation or deletion of a component, interior vertices represent the merge or split of two components and each edge illustrates a component for all isovalues between the values at each end of the edge. Note that changes in the level set and the graph only appear at critical points: $\frac{\partial f}{\partial x}(p) = 0$. Thus, those points are used to define and select meaningful level sets [6].

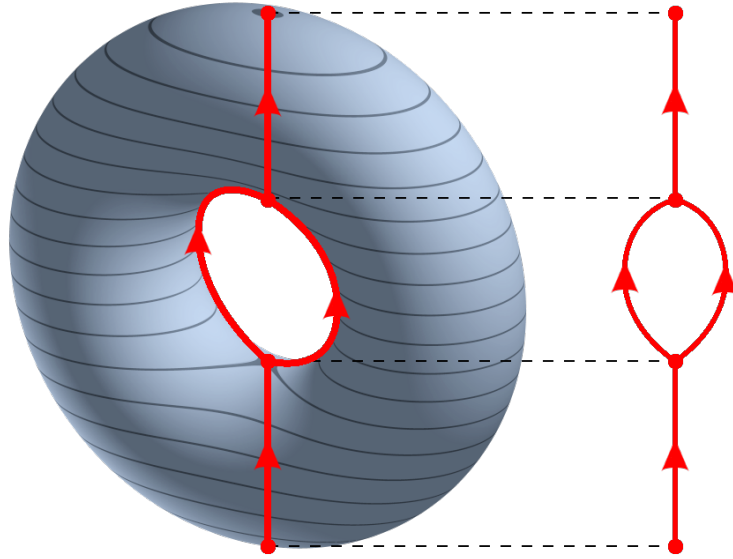


FIGURE 3.2: Reeb graph of the height function on a torus. The images are received from the Wikipedia repository [178].

If the manifold has no holes, the connected component of a level set $f^{-1}(c)$ is a closed *isocontour* and separates the domain into two regions, inside and outside of the component. This also separates the Reeb graph into two distinct subgraphs only connected by the node or edge corresponding the isocontour. Hence, the graph is a tree and is denoted as contour tree.

As an example, Figure 3.2 shows a commonly used 2D manifold in form of a torus surface with f the height function along the z -axis. Next to the torus, a Reeb graph visualizes its structure. Note that neither the Reeb graph nor the contour tree are unambiguously, i.e. two fields can have the same contour tree/Reeb graph as shown in Figure 3.3. For further literature, see for example Doraiswamy and Natarajan or Pascucci et al. [48, 128].

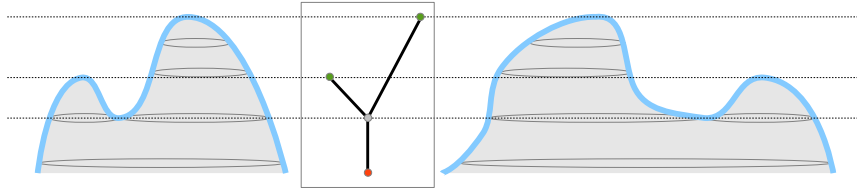


FIGURE 3.3: Two height maps and, between them, a contour tree corresponding to both maps.

3.2.3 Reeb space

Let \mathbb{M} be a d -dimensional domain, note that for a function $f : \mathbb{M} \mapsto \mathbb{R}$ the level set is $d - 1$ dimensional. The level set for a multifield $f = (f_1, \dots, f_n)$ with $f_i : \mathbb{M} \mapsto \mathbb{R}$ is defined as $f^{-1}(c) = \{x \in \mathbb{M} \mid \forall_{i=1}^n f_i(x) = c_i\}$ for each $c \in f(\mathbb{M}) \subseteq \mathbb{R}^n$. For multifields, the level set is $d - n$ dimensional, for $d \geq n$ and zero otherwise. Note that c is a n dimensional value and can change in n different directions such that the evolution of the connected components $f^{-1}(c)$ has to be illustrated in a n -dimensional space, the *Reeb space* [141]. Figure 3.4 shows an example from work by Carr and Duke [29].

A feature of both Reeb space and Reeb graph is that the structures have no geometric connection. Thus, the both visualizations are invariant to rotation, scaling or point symmetric transformation.

3.2.4 Morse-Smale Complex

For a vector field $f : \mathbb{M} \mapsto \mathbb{R}^d$, with $\mathbb{M} \subseteq \mathbb{R}^d$ some manifolds. Similar to scalar fields, critical point x is defined by $f(x) = \vec{0} = (0, \dots, 0)$. The integral line of a vector field is a path $p : (0, 1) \mapsto \mathbb{M}$ such that

- p is continuous,
- the limits/end points $p[0]$ and $p[1]$ are critical, and
- the tangent vectors of p agree with the gradients of f at all points along the path, i.e. $\forall_{x \in [0,1]} \frac{\partial}{\partial x} p(x) = \nabla f(p(s))$. [54]

Hence, integral lines represent the flow along the gradients between critical points. Note that all integral lines have several properties.

- Two lines are either the same or disjoint, i.e. do not intersect.
- Integral lines cover the entire domain \mathbb{M} .

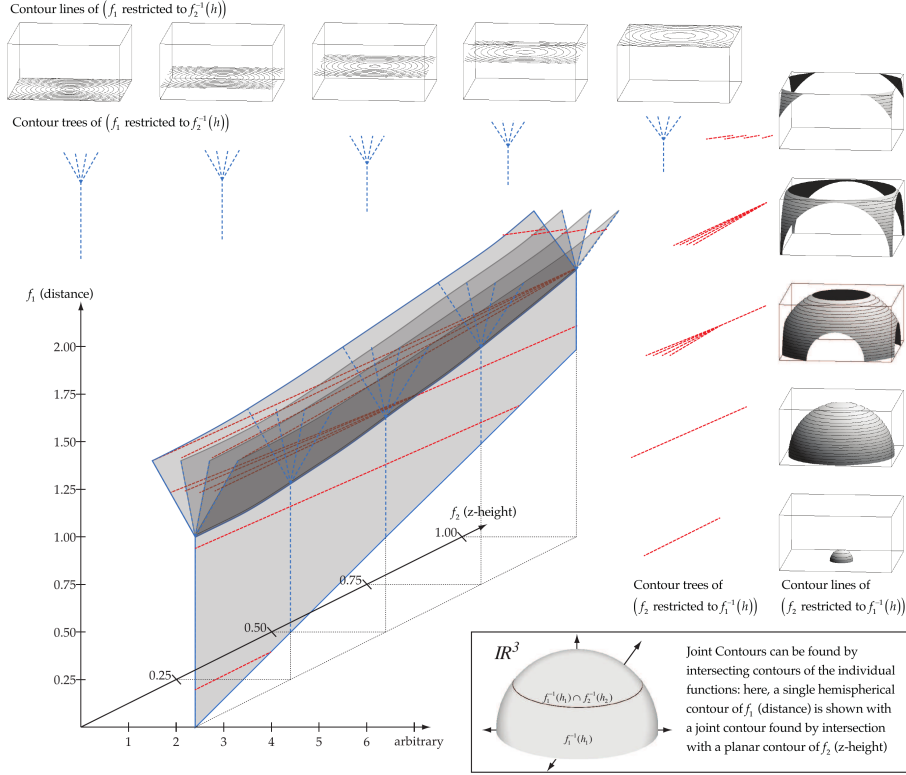


FIGURE 3.4: A illustration from Carr and Duke [29] with two 3D fields f_1 and f_2 . The field $f_1(x, y, z) = z$ is the height map, and the field $f_2(x, y, z) = \sqrt{x^2 + y^2 + z^2}$ is a distance map. The fields are further illustrated through sequences of contour trees of f_1 restricted to an isosurface for f_2 , i.e. $f_2^{-1}(h)$ for some value h , and vice versa.

With this, the domain is decomposed into *Morse-Smale cells*, regions of similar flow. For a critical point x , the *stable manifolds* $S(x)$, *unstable manifolds* $D(x)$ is given as

$$S(x) = \{x\} \cup \{y \in \mathbb{M} \mid p[1] = x \wedge p[t] = y, t \in (0, 1) \text{ for a integral line } p\},$$

$$D(x) = \{x\} \cup \{y \in \mathbb{M} \mid p[0] = x \wedge p[t] = y, t \in (0, 1) \text{ for a integral line } p\}.$$

The Morse-Smale cells are regions of points with the same destination and origin based on the integral lines as shown in Figure 3.5, alternatively a cell is defined as $S(a) \cap D(b)$ for two critical points $a, b \in \mathbb{M}$. The collection of all cell is called the *Morse-Smale complex* [76, 74, 56].

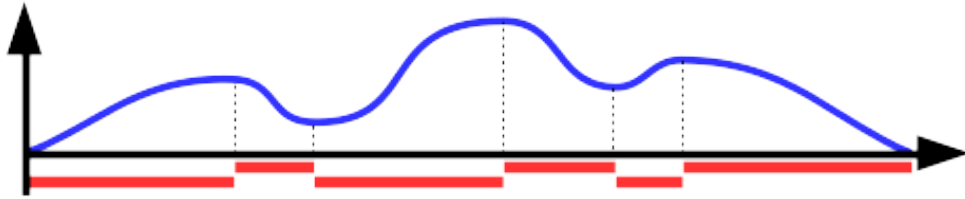


FIGURE 3.5: Example of the Morse-Smale complex for a simple 1D field. The Morse-Smale cells are indicated by the displaced red sections underneath the x-axis.

The cells can also be calculated for any scalar field f by considering the gradients of f a vector field. For multifields however, the Morse-Smale complex cannot be calculated since integral lines with tangents parallel to multiple gradient vectors usually do not exist.

3.3 Gaussian functions

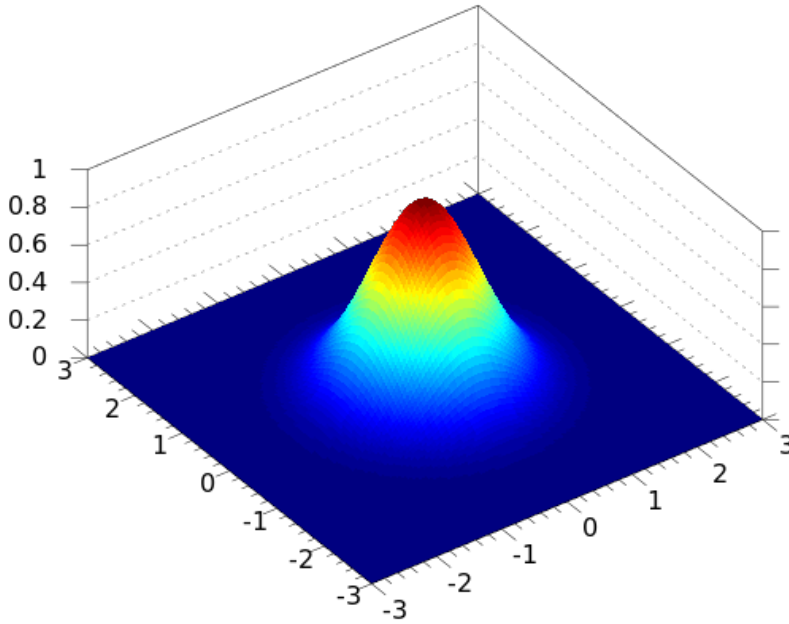


FIGURE 3.6: Gaussian curve $z = e^{-x^2-y^2}$ over a simplicial 2-complex from the free media repository at Wikipedia [179].

Throughout this work, analytical examples are used to illustrate definitions and ideas. In general, triangulated Gaussian bells of the form

$$f(x) = ae^{-\frac{(x-b)^2}{2c^2}}$$

are utilized. They create a bell curve with a as the height of the curve's peak, b the position of the center of the peak and c controlling the width of the "bell". The function values increase exponentially towards b such that f has a maximum at $x = b$. The inverse, $-f$, creates an exponentially decreasing minimum at position b . The bell curve can be calculated in any dimension, especially in 2D and 3D. Figure 3.6 shows an example on a two-dimensional plan. To create multiple peaks, i.e. minima and maxima, the Gaussian function is the sum of multiple Gaussian bells. Thus, for the implementation of a single field f the simplified version :

$$f(x) = \sum_{k=0}^m a_k e^{-2 \cdot |x-b_k|}$$

with $a_k \in \{1, -1\}$ and b_k a set of positions is used. Note that x is a critical point for f , with respect to single-field scalar topology, if and only if (iff) $x = b_k$ for some

$1 \leq k \leq m$ under the assumption that the distance $|b_k - b_{k'}|$ for all $1 \leq k, k' \leq m$ is large enough such that $\forall_{k \neq p} e^{-2 \cdot |x - b_k|} \ll \epsilon$ for all $x \approx b_{i,p}$ and some small $\epsilon > 0$ can be assumed. Note that this is not a impossible assumption since $e^{-2 \cdot |x - b_k|}$ converges exponentially.

The scalar a_k allows to define the type of critical point at position b_k , in detail a minimum if $a_{i,k} = -1$ and a maximum if $a_{i,k} = 1$. Saddle points, however, cannot be defined directly and do not correspond to any $b_{i,k}$.

Throughout the chapters, if an example field is created with this concept, the multifield is only based on rather similar Gaussian bells. Therefore, m , the number of positions $b_{i,k}$, is equal in all fields f_i and furthermore, the positions are only shifted slightly over the individual fields, i.e. $|b_{i,k} - b_{j,k}| < \epsilon$ for all $1 \leq i, j \leq n$ and some $\epsilon > 0$. The position distance $|b_{i,g} - b_{i,h}|$ inside each field f_i , however, can be arbitrary. Hence, to create a multifield with n fields, it is sufficient to define a matrix of $n \times m$ positions.

Chapter 4

Definition and Computation of the Pareto Set

4.1 Motivation

The concepts that are presented in the following chapters are aimed at elucidating topological structures of multiple fields. While its implementation assumes piecewise linear fields over simplicial complexes, the definition only requires Morse fields over the same manifold \mathbb{M} . Alternative definitions are also provided to use in later chapters and inspire alternative implementations.

With the focus on visualization, application, and implementation, the definition is shifted in a later part of this chapter to simplicial complexes and piecewise linear Morse fields, especially for purposes of application and simplification of Pareto sets in Chapter 5 and 7.

Furthermore, this chapter is the collection of all corresponding definition sections in previously published papers and conference presentations [91, 86].

4.2 Definition of Pareto Optimality

In this sections ideas and concepts pioneered in the domain of multi-criteria optimization are described and adapted to derive a topological-based description of a multifield f . The presentation follows Stadler and Flamm [152], who describe Pareto sets on graphs, and adapts their description to the continuous setting of multifield $f : \mathbb{M} \mapsto \mathbb{R}^n$, $\mathbb{M} \subseteq \mathbb{R}^d$ with an adjusted terminology. Note that f can be described as a set of n individual single-fields $f_i : \mathbb{M} \mapsto \mathbb{R}$ such that $f = (f_1, \dots, f_n)$.

Multi-criteria Optimization Stadler and Flamm adapted definitions from Ehrgott and Gandibleux [59]. In their context, a (multiobjective) combinatorial problem is defined as $X \subseteq 2^A$, with A a finite set of configurations, several weight functions $w_j : A \mapsto \mathbb{Z}$, yielding a finite set of objective functions f_i . Two usual examples, given by Ehrgott and Gandibleux are

$$f_i(S) = \sum_{a \in S} w_i(a), \text{ or}$$

$$f_i(S) = \max_{a \in S} w_i(a),$$

with $S \in X$. The problem is solved by identifying

$$"min"_{S \in X}(f_1(S), \dots, f_n(S))$$

for some meaning of "min". In other words, optimization aims to find a configuration $S \in X$ such that no other configuration $T \in X$ is better than S . Better in terms of Stadler and Flamm means, that $f_i(S) \leq f_i(T)$ for all i , who defined the optimal solutions as Pareto minima.

In this thesis and several other works [43, 169, 150], this definition is extended from the finite configuration space to continuous manifolds.

To compare two different points x and y in \mathbb{M} with respect to their corresponding values of the multifield field f , it is said that x (weakly) dominates y if $f_i(x) \geq f_i(y)$ for all $i = 1, \dots, n$, and denoted as $x \succeq y$. Furthermore, x strictly dominates y , written as $x \succ y$, if $x \succeq y$ and there is at least one index i such that $f_i(x) > f_i(y)$. x and y are called comparable if either $x \succeq y$ or $y \succeq x$ holds, and incomparable ($x \prec \succ y$) otherwise. Intuitively, comparability is based on the notion that all fields f_i change in the same way – all either increase or decrease – when moving from x to y .

Note that the relation \succeq defines a partial order on the values of f :

Reflexivity Given a point $x \in \mathbb{M}$, it is obvious that $f_i(x) \leq f_i(x)$ for all $1 \leq i \leq n$. Hence $x \succeq x$ holds.

Transitivity Given three points $x, y, z \in \mathbb{M}$ with $x \succeq y$ and $y \succeq z$, this implies that for all $1 \leq i \leq n$, $f_i(x) \geq f_i(y)$ and $f_i(y) \geq f_i(z)$ and therefore also $f_i(x) \geq f_i(z)$. Hence $x \succeq z$ holds.

Antisymmetry Given two points $x, y \in \mathbb{M}$ such that $x \succeq y$ and $y \succeq x$ hold. Hence, for each individual fields f_i it holds that $f_i(x) \geq f_i(y)$ and also $f_i(x) \leq f_i(y)$. Since all f_i are Morse fields, this implies that $f_i(x) = f_i(y)$ and thus $f(x) = f(y)$.

In contrast to Stadler and Flamm, the manifold domain \mathbb{M} is used to define a open neighborhood $U(x)$, a connected area of \mathbb{M} without borders and containing $x \in \mathbb{M}$. This supports later visualization approaches by not only identifying global, but also local critical points.

Definition 4.2.1. In this notation, a point x is called a (local) Pareto optimum if there exists an open neighborhood $U(x)$ containing x such that

$$\forall U'(x) \subseteq U(x), \forall y \in U(x) \setminus \{x\}, x \prec \succ y.$$

Similarly, x is a (local) Pareto minimum if it has comparable points $y \in U(x) \setminus \{x\}$, but is weakly dominated by them, i.e.

$$\forall U'(x) \subseteq U(x), \forall y \in U(x) \setminus \{x\}, (x \prec \succ y) \vee (x \succeq y).$$

If, on the other side, x weakly dominates all comparable points in the neighborhood, such that

$$\forall U'(x) \subseteq U(x), \forall y \in U(x) \setminus \{x\}, (x \prec \succ y) \vee (x \preceq y)$$

holds the point x is defined as a (local) Pareto maximum. All other points are termed regular.

For the remainder of this work, Pareto optima, Pareto minima, and Pareto maxima are collectively referred to *Pareto extrema*. Thus, this thesis defines the sets

$\mathbb{P}^+(f)$ of all Pareto maxima,

$\mathbb{P}^-(f)$ of all Pareto minima,

$\mathbb{P}^o(f)$ of all Pareto optima, and

$\mathbb{P}(f)$ of all Pareto extrema

for the multifield f , with $\mathbb{P}(f)$ also referenced as the *Pareto set*. If the field f is clear from the context, it can be neglected, i.e. \mathbb{P} is sufficient.

Obviously, $\mathbb{P}^+(f) \cup \mathbb{P}^-(f) \cup \mathbb{P}^o(f) = \mathbb{P}(f)$ holds directly implied by their definition. Note that in contrast to existing work [43, 169, 150] this thesis distinguishes between three types of Pareto extrema instead of just Pareto optimality to consider visualization and application tasks in current and future work.

The definitions are illustrated in Figure 4.1, which shows all types of points in the Pareto set concept with two 1D fields. See, for example, the location x marked as Pareto minimum. Both function g and h increase towards the left, while to the right g decreases and h increases. Hence, to the right, the x is adjacent to incomparable points and, to the left, x is adjacent to dominating points such that x is defined a Pareto minimum.

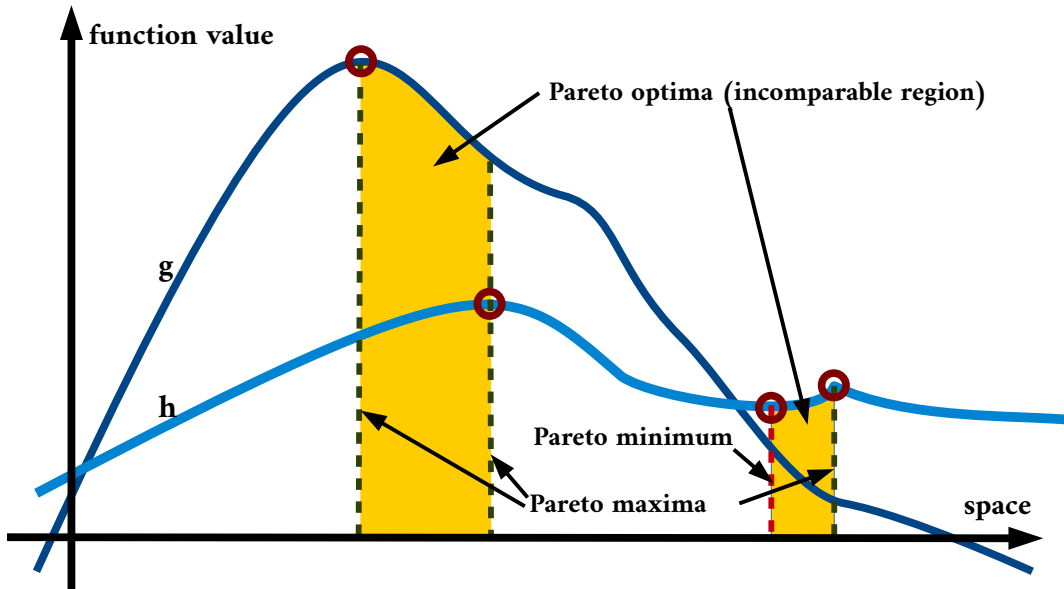


FIGURE 4.1: Two fields g, h with individual critical points (red circles). The location of Pareto maxima and Pareto minima are indicated by green and red dotted lines, respectively. Pareto optima create two connected areas, colored yellow.

It should be noted that in the single-field case ($n = 1$), minima/maxima are Pareto minima/maxima according to this definition. Thus, in this regard, the approach is analogous to scalar single-field topology. Furthermore, for $n > 1$, each

point that is maximal or minimal in at least one individual field is either a Pareto maximum or Pareto minimum, respectively, or a Pareto optimum. In contrast however, other critical points, i.e. so-call saddle points, are not directly translated to Pareto extrema and remain an open task.

To calculate the Pareto set, i.e. classify points as Pareto extremal or regular points, it is necessary to identify the comparable points with respect to these points. Comparable points with respect to a point x are separated into those dominating x and those dominated by x , resulting in the following definitions.

Definition 4.2.2. *Given any open neighborhood $U(x)$ with $x \in U(x)$, for each field f_i , the set*

$$H_{i,U(x)}^+(x) := \{ y \in U(x) \mid f_i(x) \leq f_i(y) \}$$

is defined. Considering all f_i , x is dominated by the local ascending set

$$H_{U(x)}^+(x) := \bigcap_{i=1}^n H_{i,U(x)}^+(x).$$

The local descending set $H_{U(x)}^-(x)$ for point x and neighborhood $U(x)$ is defined analogously, and let $H_{U(x)}^\pm(x) := H_{U(x)}^+(x) \cup H_{U(x)}^-(x)$ is the set of all comparable points around x restricted by $U(x)$.

Note that H^+ and H^- can be calculated iteratively with H_i^+ and H_i^- , respectively, and can be used as an alternative definition of the Pareto set. This equality is directly implied by the corresponding definitions and yields the following lemma.

Lemma 4.2.1. *The point $x \in \mathbb{M}$ is a Pareto optimum, iff there exists a $U(x)$ such that*

$$H_{U'(x)}^\pm(x) := \{x\} \forall U'(x) \subseteq U(x).$$

The analogous lemma for H^+ and Pareto maxima as well as H^- and Pareto minima are also direct implication of their definitions. The restriction to consider all subsets $U'(x) \subseteq U(x)$ can be dropped, if f is linear in $U(x)$.

Proof. The multifield f is linear, iff every individual field f_i is linear. Hence each f_i is of the form $g(x) = a + b \cdot x$ for some scalar $a \in \mathbb{R}^d$ and $b \in \mathbb{R}$. Without loss of generality, it can be assumed that $a = \vec{0}$. Let $y \in H^+_{i,U(x)}(x)$ exist with $y \neq x$ such that $f_i(x) = g(x) \leq g(y) = f_i(y)$.

For every subset $U'(x) \subseteq U(x)$, a point z exists such that $z = (1-t) \cdot x + t \cdot y$ for some $t \in (0, 1)$, since $U'(x)$ is open. For this point z the linear function g implies that

$$f_i(z) = g(z) = (1-t) \cdot g(x) + t \cdot g(y) \geq (1-t) \cdot g(x) + t \cdot g(x) = g(x)$$

and $z \in H^+_{i,U'(x)}(x)$ hold. This argument can be repeated for every individual field f_i such that $z \in H^+_{U'(x)}(x)$ and thus $H^+_{U'(x)}(x) \neq \{x\}$ is proven. \square

With respect to the focus of this thesis, an alternatively definition of H^+ arises under the assumption that f_i are linear inside a neighborhood $U(x)$ and each field can be written in the form $g(x) = a + b \cdot x$, or more specifically in the form $g_i(x) =$

$f_i(0) + \nabla f_i(x) \cdot x$. Note that for these linear fields ∇f_i is constant over all $x \in U(x)$. Hence, for two points $x, y \in U(x)$ and every $1 \leq i \leq n$ it holds that

$$f_i(x) \leq f_i(y) \Leftrightarrow g_i(x) \leq g_i(y) \Leftrightarrow \nabla f_i(x) \cdot x \leq \nabla f_i(x) \cdot y$$

and therefore

$$f_i(x) \leq f_i(y) \Leftrightarrow 0 \leq \nabla f_i(x) \cdot (y - x).$$

As a direct implication, the local ascending set can be defined over the gradients $\nabla f_i(x)$ for the individual fields f_i . In other words, for two points x, y , the inequality $f_i(x) \leq f_i(y)$ holds if y is contained in the set

$$H_{i,U(x)}^+(x) := \{ y \in U(x) \mid (y - x) \cdot \nabla f_i(x) \geq 0 \}.$$

Also the local descending set can be alternatively defined through the term $(y - x) \cdot \nabla f_i(x) \leq 0$. This alternative definition allows for an alternative implementation, as shown later, and opens the Pareto set concept to be understood from different points of view.

Intuitively, Pareto extremal points represent barriers across which the components of the multifield function f cannot be jointly increased or decreased further. In other words, it is not possible to strictly increase one or more components while keeping the others equal. See for example Figure 4.1, where the barriers are clearly visible through the yellow Pareto optimal regions. It is therefore motivated to join adjacent points that are Pareto extrema into connected components, thus a barrier. More specifically, two points are called connected if their are adjacent in the domain. This relation is transitive, such that its reasonable to identify maximally connected subsets among the Pareto extrema as *connected components* in the Pareto set \mathbb{P} . For reference in following parts of this work, $\mathbb{E}(f)$ denotes the set of connected components in $\mathbb{P}(f)$.

4.3 Global Ascending and Descending Sets

To understand the global structure of the multifield f , an intuitive notion is derived from paths along which all individual fields increase or decrease, and connect minimal and maximal structures via such paths, in rough analogy to vector field topology.

Definition 4.3.1. To achieve this, an ascending path $p^+ : [0, 1] \rightarrow \mathcal{S}$ is defined as a continuous map from the unit interval to \mathcal{S} with the property that

$$f(p(s)) \preceq f(p(t)) \quad \text{for all } 0 \leq s \leq t \leq 1.$$

Intuitively, as one traverses p^+ , all individual fields f_i do not decrease. This definition implicitly requires that any two points on an ascending path are comparable. If two points x, y are connected by some ascending path p such that $p(0) = x$ and $p(1) = y$, in the remainder of the thesis this circumstance is denoted as $x \rightsquigarrow y$.

Definition 4.3.2. Using ascending paths, the global ascending set of a point is defined as the set of all points in \mathcal{S} which can be reached by an ascending path from x , i.e.

$$\mathcal{C}^+(x) := \{ y \mid \exists p^+ \text{ s.t. } p^+(0) = x \wedge p^+(1) = y \} = \{ y \mid x \rightsquigarrow y \}.$$

Obviously, *descending paths* and *global descending sets* can be defined in complete analogy. Note the relation between the global and local ascending set: $H_{U(x)}^+(x) = \mathcal{C}^+(x) \cap U(x)$ for a sufficiently small neighborhood $U(x)$, but not in general since an ascending path can leave $U(x)$, loop around and reenter the neighborhood, thus containing both points in- and outside of $H_{U(x)}^+(x)$. Therefore, in this work different notations for global and local ascending sets are used. Figure 4.2 illustrates the difference in a small example on a 3×3 data point set with two sets of values. Both images show the neighborhood of the Pareto minimum $(0, 0)$ as hatched area. Hence, the local ascending set contains the points $(1, 1)$ and $(6, 6)$. The global ascending set is implied by the path (black lines) originating from $(0, 0)$. Note that, among others points, $(6, 6)$ is only contained in the global ascending set in the first image (a) and not in the second (b) due to the barrier of Pareto optima.

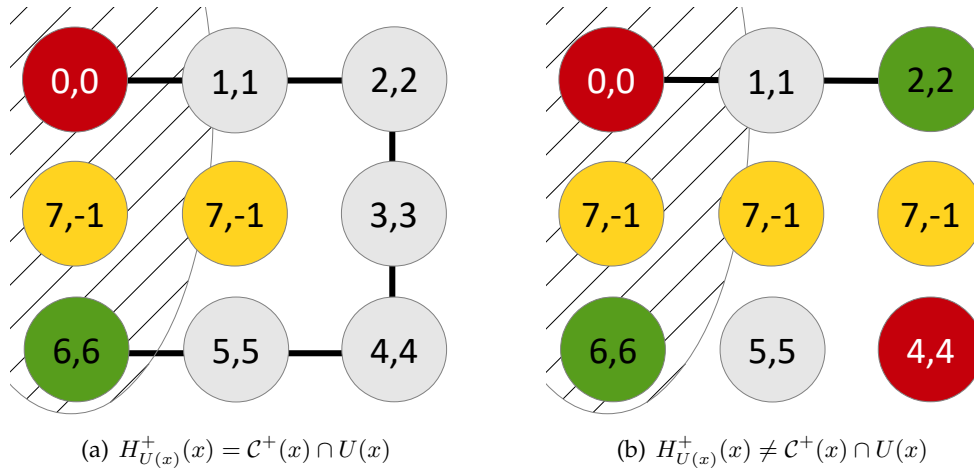


FIGURE 4.2: 3×3 data points with two sets of values. Pareto maxima, Pareto minima, and Pareto optima are colored green, red, and yellow, respectively. A neighborhood around the point with values $(0, 0)$ is indicated by a hatched area. Ascending paths from this points are indicated by a black line.

Furthermore, an ascending path connecting two points x and y , i.e. $x \rightsquigarrow y$, is not necessarily unique. Consider for example the 2D-function $f(a, b) = a$ and the two points $x = (0, 0)$ and $y = (1, 1)$. Any path from x to y with an monotone increasing a coordinate, regardless of the b coordinate, is an ascending path, for example $(0, 0) \rightarrow (0, 1) \rightarrow (1, 1)$ and also $(0, 0) \rightarrow (1, 0) \rightarrow (1, 1)$. Hence, it is beneficial to think of ascending and descending paths as extensions of the notion of monotone paths for single scalar fields in contrast to steepest gradient paths. Note that ascending and descending sets behave like mirrors: If a point x is inside the ascending set of a point y then y is in the descending set of x .

Lemma 4.3.1. *To show that local and global ascending sets are interchangeable for the definition of Pareto maxima, the following proof shows that for every point $x \in \mathbb{M}$*

$$(\exists_{U(x)} \forall_{U'(x) \subseteq U(x)} H_{U'(x)}^+(x) = \{x\}) \Leftrightarrow (\mathcal{C}^+(x) = \{x\}).$$

Proof. Let $y \in C^+(x)$ be a point with $x \neq y$ and $x \rightsquigarrow y$ such that an ascending path p^+ exists with $p^+(0) = x$ and $p^+(1) = y$. Let $U(x)$ be any open neighborhood around x that contains both x and y . For every subset $U'(x) \subseteq U(x)$, it is possible to find $j \in (0, 1]$, the smallest position such that $\forall_{i \in [0, j]} p[i] \in U'(x)$, since $x \in U'(x)$ and $U'(x)$ is open. Hence $p^+(j) \in H_{U'(x)}^+(x)$ with $p^+(j) \neq x$ exists.

Let $U(x)$ a neighborhood around x such that $\forall_{U'(x) \subseteq U(x)} H_{U'(x)}^+(x) \neq \{x\}$. Assume that some $y \in U(x)$ is a point with $y \succeq x$ and $x \neq y$ but without any ascending path p^+ between x and y . In other words, for every continuous path $p : [0, 1] \mapsto \mathbb{M}$ with $p[0] = x$ and $p[1] = y$ some position $a, b \in [0, 1]$, $a < b$, exists such that $f(p(a)) \prec f(p(b))$ or $f(p(a)) \prec \succ f(p(b))$ holds. For any of these paths p , let $j \in (0, 1]$ be the smallest such positions. Note that p and j have to exist since otherwise x is surrounded by incomparable points, i.e. a Pareto maximum or f is not a smooth Morse function. Therefore, the section of the path from $p[0]$ to $p(j)$ define an ascending path $q(i) = p(i/j)$ that implies $x \rightsquigarrow p(j) \neq x$, $p(j) \in C^+(x)$ and thus $C^+(x) \neq \{x\}$. \square

Taking the union of ascending (resp. descending) paths over all points in a small neighborhood and further over a set of points, allows the straightforward generalization of ascending and descending sets for Pareto extremal regions, i.e. elements of $\mathbb{E}(f)$. To fix terminology, let $P \in \mathbb{E}(f)$ be a (connected) Pareto minimal region, thus a union of multiple, Pareto minimal points, then $\mathcal{C}^+(P) := \bigcup_{x \in P} \mathcal{C}^+(x)$ is the (global) ascending set of P . Again, (global) descending sets are defined in complete analogy.

4.4 Definition of the Reachability Graph

With the notation of connected components in $\mathbb{P}(f)$ and global ascending/descending sets, in this thesis an abstract view on the structure within the multifield f . The structure represents the components $\mathbb{E}(f)$ and how they are connected by ascending and descending paths. For the remaining chapters, component $R1$ reaches component $R2$ ($R1 \rightsquigarrow R2$), iff there exists a point $x \in R2$ such that $x \in \mathcal{C}^+(R1)$.

Definition 4.4.1. *Given the global ascending and descending sets for each connected component in $\mathbb{P}(f)$, the reachability graph $RG(f) = (V, E)$ for f is constructed. Each component is translated to a node in the graph, i.e. a bijection $b : \mathbb{E}(f) \mapsto V$ exists. For edges, the graph contains a directed edge from nodes $b(R1)$ to $b(R2)$, iff $R1$ reaches $R2$. In other words,*

$$\forall R1, R2 \in \mathbb{E}(f), (R1 \rightsquigarrow R2) \Leftrightarrow ((b(R1), b(R2)) \in E).$$

Figure 4.3 sketches a scenario with four fields, each based on a sum of Gaussian bell fields and each with a uni-single maximum in $R2, R4$, a minimum in $R3, R4$, and a saddle in $R1$. The ascending and descending sets of $R1$ are colored in red and green, the descending set for $R2$ is hatched to exemplarily indicate reachability. The second image shows the reachability graph for this scenario. As a reminder, Pareto extremal points represent barriers, such that ascending paths cannot move through Pareto optimal regions, i.e. regions of incomparable points, or through Pareto maxima or Pareto minima. Hence, reachability among the connected components is not transitive.

The general concept is able to classify points as Pareto maxima, Pareto minima, and Pareto optima while the elements in $\mathbb{E}(f)$ can only be denoted as Pareto extrema

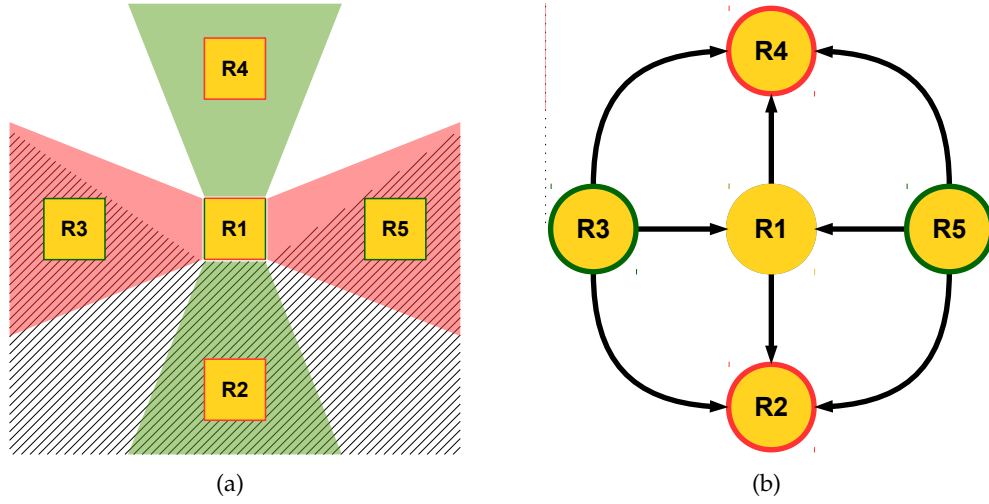


FIGURE 4.3: An example configuration of Pareto extrema illustrating the reachability graph, the global ascending and descending sets of the connected component $R1$ (shaded red and green), as well as the descending sets of $R2$ (hatched).

since they can also consist of a mix of these, see for example $R1$ in Figure 4.3 or the connected component of Pareto extrema in Figure 4.1. Through the number of outgoing and incoming edges in the graph it is possible to give a refined classification of the elements in $\mathbb{E}(f)$. Between *Pareto maxima* and *Pareto minima* can be distinguished, if only ingoing, respectively only outgoing, edges exist at the corresponding node in the reachability graph. If for a node incoming and outgoing edges exist, the corresponding element in $\mathbb{E}(f)$ is denoted as a *Pareto saddle*. Again, note the relation to the terminology of Reeb graphs, where nodes with only outgoing edges correspond to minima, nodes with only ingoing edges correspond to maxima and all other nodes correspond to saddle points [127, 32].

Also note that loops in the graph are possible. A current assumption is, that some of these are based on numerical errors resulting from the triangulation of the domain / the field. As a remark, other concepts like Jacobi sets seem to have similar problems [113].

For a simple example, the input data as presented in Figure 4.4 is presented, a ring of four vertices $\{v_1, v_2, v_3, v_4\}$ connected by edges (v_1, v_2) , (v_2, v_3) , (v_3, v_4) , and (v_4, v_1) . The field values of two fields are printed directly on the vertices while all other points along the edges are linearly interpolated. The field values at the vertices are: $f(v_1) = (1, 1)$, $f(v_2) = (2, 2)$, $f(v_3) = (3, 3)$, and $f(v_4) = (0, 4)$. Note that v_1 and v_4 , and also v_2 and v_4 are incomparable such that these vertices and all points included in the two edges between them are Pareto extremal and form a connected component in $\mathbb{P}(f)$. However, v_1 reaches v_3 via a path over v_2 such that the resulting reachability graph contains only one node and a loop to itself.

Hence, it is necessary to refine the classification to ignore loops in the number of ingoing and outgoing edges to define Pareto minimal, Pareto maximal, and Pareto optimal elements of $\mathbb{E}(f)$.

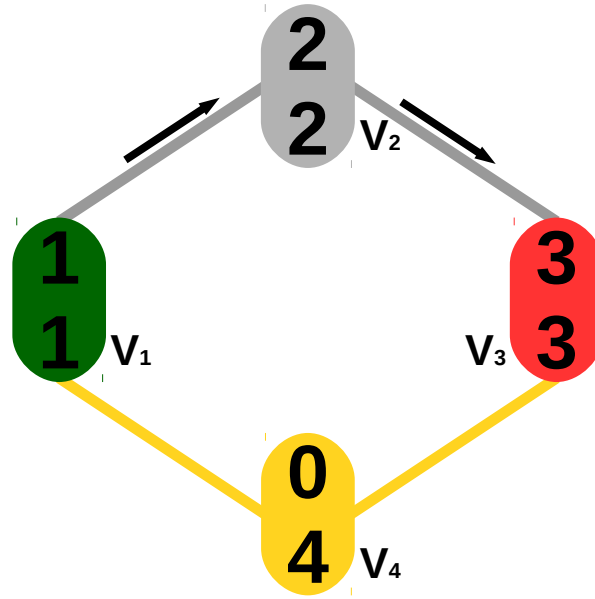


FIGURE 4.4: A simple example with two sets of scalar values over four vertices and four edges. Vertex v_1 is a Pareto minimum, colored in green, v_3 is Pareto maximal, colored in red, and vertex v_4 and adjacent edges, colored in yellow, are Pareto optimal. v_1, v_3 , and v_4 and all points along the mentioned edges build on connected component S , i.e. one element in $\mathbb{E}(f)$. Note that this component reaches itself via the ascending path along the edges (v_1, v_2) and (v_2, v_3) which results in a loop.

Definition 4.4.2. Let $RG = (V, E)$ be the reachability graph for some multifield f and $v \in V$ some node in the graph. If the number of outgoing edges

$$out = |\{u \in V \mid (v, u) \in E \wedge u \neq v\}|$$

is zero ($out = 0$), the connected component in $\mathbb{E}(f)$ corresponding to v is called a Pareto maximum. If the number of ingoing edges

$$in = |\{u \in V \mid (u, v) \in E \wedge u \neq v\}|$$

is zero ($in = 0$), the component is a Pareto minimum. If neither holds, the component is called a Pareto saddle.

Additionally, note that an edge (g, h) in the reachability graph implies that the connected component G reaches H , i.e. that

$$\mathcal{C}_{G,H} := \{x \in \mathbb{S} \mid x \in \mathcal{C}^+(G) \wedge x \in \mathcal{C}^-(H)\} \neq \emptyset$$

holds, but it is possible that this set $\mathcal{C}_{G,H}$ consists of multiple, separate connected components. Hence, while not in the focus of this section, a multigraph with multiple edges between the nodes might be a more suited visualization. This and other aspects, like for example a good placement of the edges according to $\mathcal{C}_{G,H}$ is a possible topic of future work, while for the simplification approach described in Chapter 5 the graph as presented above is sufficient.

4.5 Implementation for Piecewise Linear Multifields

In this section, implementation guidelines are provided to identify Pareto extremal points in pseudocode. For the result described in later chapters, the programming language $C++$ [155] is used to calculate the Pareto set, its related structures and visualizations, but the pseudocode is flexible enough such that an alternative implementation in another languages is feasible.

Furthermore, for the purpose of implementing an algorithm that (de)classifies points as Pareto extrema and calculates their ascending and descending sets, the domain of the general concept is changed from continuous fields to piecewise linear scalar fields given on the same simplicial complex \mathbb{S} of dimension d . In general, the scalar values of the n fields are only given at the vertices and stretched over the whole domain \mathbb{S} through bilinear interpolation (see the corresponding section in Chapter 3).

As a direct implication, if $x, y \in \mathbb{S}$ are two points inside the same simplex σ and do not lie on its faces / border, the set of comparable points with respect to x is related to the set of comparable points with respect to y simply by a translation.

Lemma 4.5.1. *If any point inside a simplex σ is Pareto maximal, minimal, or optimal, all points $y \in \sigma$ are Pareto maximal, minimal, or optimal, respectively.*

Proof. Let $x \in \sigma$ be Pareto minimal, thus in a small neighborhood $U_\epsilon(x) := \{u \in \sigma \mid |x - u| < \epsilon\} \subseteq \sigma$ around x all points are either incomparable or dominating x . Note that this also holds for all subsets $U_{\epsilon'}(x) \subseteq U_\epsilon(x)$, for any $\epsilon' < \epsilon$. Let $y \in \sigma$ be any other point inside the simplex with a neighborhood $U_{\epsilon''}(y) \subseteq \sigma$ with $\epsilon'' < \epsilon$.

Since all fields f_i are linear inside σ , specially in $U_{\epsilon''}(y)$ and $U_{\epsilon''}(x)$, they can be approximated with $f_i(x) = a_i \cdot x + b_i$ for some variables $a_i, b_i \in \mathbb{R}, 1 \leq i \leq m$.

Let $x' \in U_{\epsilon''}(x)$ be any point near x , i.e. $x' = x + z$ for some $|z| < \epsilon''$, for which, w.l.o.g. $f_i(x) < f_i(x')$ holds.

$$f_i(x) < f_i(x') \Rightarrow a_i \cdot x + b_i < a_i \cdot x' + b_i = a_i \cdot (x + z) + b_i.$$

Hence,

$$a_i \cdot x + b_i < a_i \cdot (x + z) + b_i \Rightarrow 0 < a_i \cdot z,$$

and thus,

$$f_i(y) = a_i \cdot y + b_i < a_i \cdot (y + z) + b_i.$$

This can be shown analogously for every other field f_i as well as for cases with $f_i(x) > f_i(x')$. Hence, if either $x \succ x'$, $x \prec x'$ or $x \prec \succ x'$ in some neighborhood $U_{\epsilon''}(x)$ holds, for any $y \in \sigma$ a $y' \in U_{\epsilon''}(y)$ is found such that $y \succ y'$, $y \prec y'$, or $y \prec \succ y'$, respectively, holds. \square

As a direct implication, the ascending and descending sets of x and y are related by translation and restriction to σ . As illustration, see how the ascending sets for x and x' in Subfigure 4.5(a) are of the same shape and size, if they would not be restricted to the triangle. This does not hold if x or y lie on a face, in other words in

a simplex of dimension $d' < d$. Then, x is located inside multiple simplices σ_j and the ascending set $H^+(x)$ is the union of multiple local ascending sets:

$$H^+(x) := \bigcup_j H_{\sigma_j}^+(x).$$

The images in Figure 4.5 illustrates this behavior inside 2D simplices.

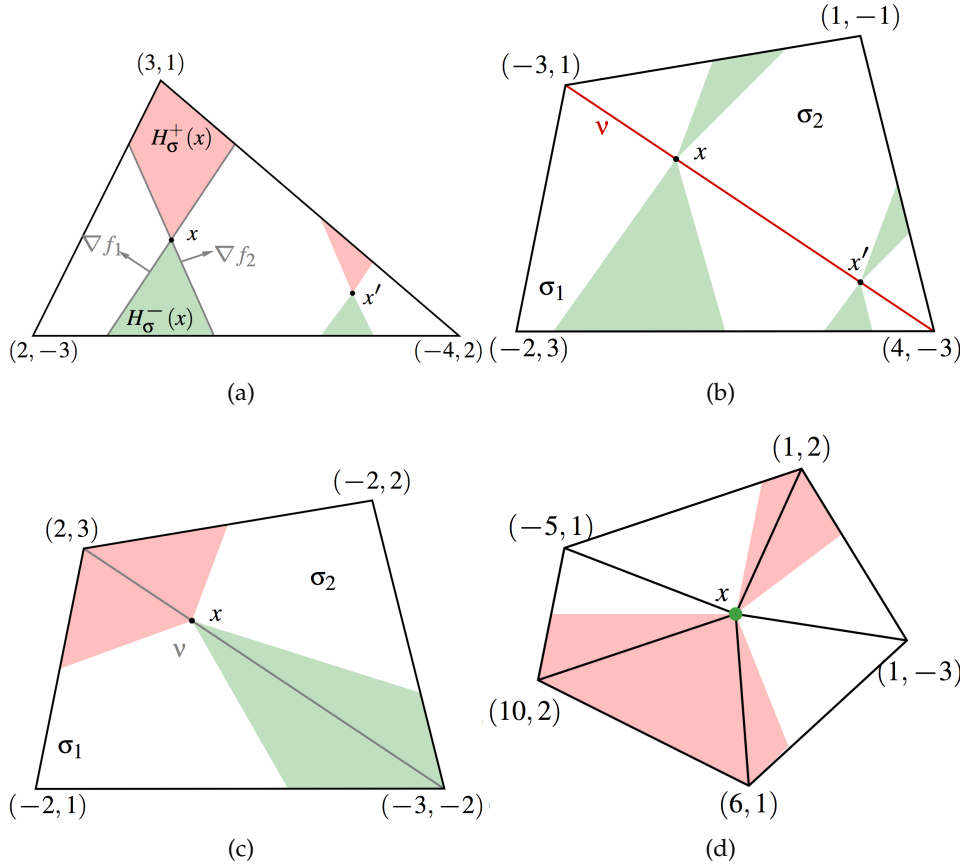


FIGURE 4.5: Different configurations with two fields over a selection of triangles to illustrate the concept of ascending (red) and descending sets (green).

Hence, to determine Pareto extremality, the implementation outlined in Listing 4.1 starts by checking for each simplex whether its barycenter is a Pareto extremum. As proven above, iff one point in the interior is Pareto extremal, all interior points are Pareto extremal and of the same type.

The algorithm tests each d -simplex $\sigma \in \mathbb{S}$ by computing the ascending set $H_{\sigma}^+(x)$ that is the intersection of the m half-spaces $H_{\sigma,i}^+(x)$ at the barycenter x and the $d + 1$ half-spaces given by σ 's faces. Note that the result is a polygon with at most m sides and x as one of its corners. Inside σ , this polygon can be considered a cone starting at x such that the ascending and descending sets can also be addressed as *ascending* and *descending cones*, respectively.

The implementation starts with the simplex σ and successively clips it by the

half-spaces that originate from non-zero gradients. This process is illustrated in Figure 4.6. In this image, a triangle is reduced by three half-space, indicated by three thick lines. The remaining area is colored in green and its representation as a simplicial complex is indicated by thinner lines.

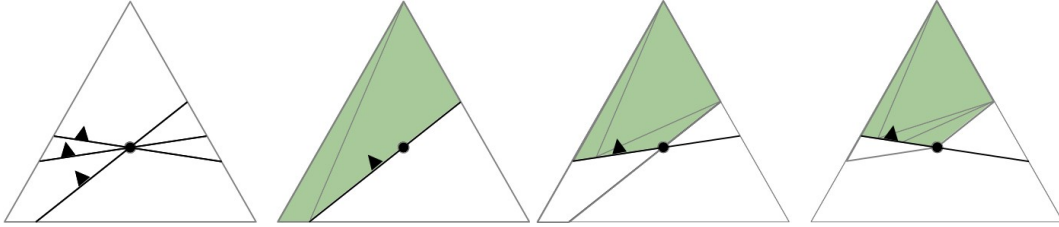


FIGURE 4.6: Successive clipping of a triangle by half-planes originating from three gradients at the barycenter to determine $H_\sigma^+(x)$ (green region).

Because, in general, a clipped simplex is no simplex but a polyhedron that can always be represented by the union of at most d simplices, a list of simplices is stored instead of a polyhedron as the clipping result. As each additional half-space intersection potentially doubles the number of simplices in this list, the worst-case runtime is in $O(d^k)$.

For 2D and 3D the *Marching Triangle* and *Marching Tetrahedra* methods can be used [99]. Note that for this implementation only the field values at the barycenter and the vertices are needed and not their domain coordinates.

The function *hasCone* determines if the barycenter with field values v has an ascending set inside a triangles given by its three field values vi at the corners. The values are implemented as n -vectors, with n the number of separate fields. If the boolean flag *ascending* is set *false*, the descending set is calculated instead. In this pseudocode, each individual field f_i is considered recursively and the current triangle is split into subtriangles containing only points with values larger than $v[i]$. The subtriangles are encoded in the set denoted *sub_tri* in the code. Each element of *sub_tri* contains first the number of new subtriangles and a sequence of triples indicating the new vertices of the each subtriangle. Each triple consists of three indices. If an index t is > 3 , the new vertex has the old values $vi[t-3]$. Otherwise, the field intersects the triangle between the vertices of t and $t + 1$ modulo 3. The intersection and thus the new values are calculated implicitly.

To choose the element of *sub_tri* that encodes the intersection of the triangle with $f_i(v)$, note that every intersection of a triangle corresponds to one of the $2^3 = 8$ classes, depending which of the three vertices is inside or outside the remaining part of the triangle. Hence, each class can be described by a 3-bit vector. Figure 4.7 shows examples of these eight classes.

If at the end all subtriangles are empty, *hasCone* returns false. Otherwise, if at least one subtriangle is non-empty, thus a set of points exists where all field values are larger than those at the barycenter, a boolean *true* is sent back through the recursion tree to be the output of *hasCone*.

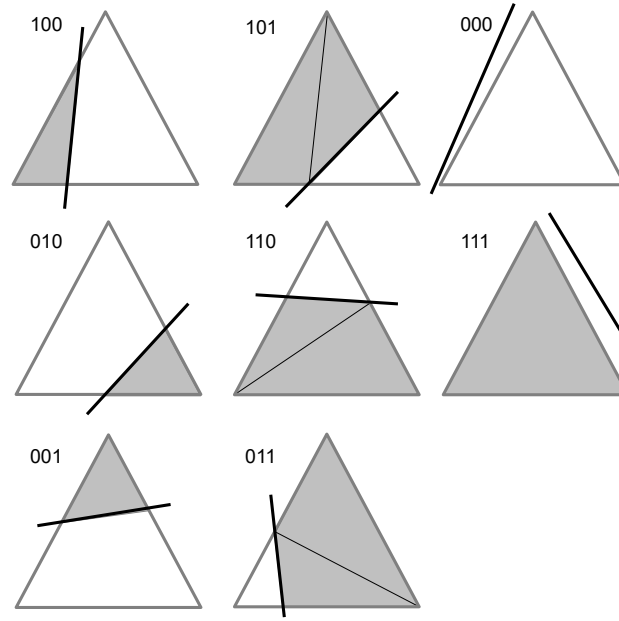


FIGURE 4.7: The 8 classes of possible triangle intersections with their corresponding 3-bit descriptor. The thick black line marks the intersection, the gray area marks the remaining triangle part and the thin black line indicates possible subtriangles.

As a reminder, since the fields are piecewise linear inside the simplices, it is sufficient for each simplex to run *hasCone* only for the field values at its barycenter as input v and, as input vi , the corresponding information from each adjacent triangle (highest dimensional simplex). These calculations can easily be multi-threaded.

LISTING 4.1: Recursive pseudocode for the ascending or descending set of a point in a triangle.

```

1  int hasCone( values v, values vi[3],
3      bool ascending, int n )
4  {
5      // all fields are considered
6      if( n >= v.size() )
7      {
8          // does the current triangle exist?
9          if( (|vi[0]-vi[1]| < 0 or
10             (|vi[1]-vi[2]| < 0 or
11             (|vi[2]-vi[0]| < 0 )
12             return false;
13
14         return true;
15     }
16
17     int sub_tri[8][7] = {
18         { 0, }, { 1, 0, 2, 3, }, { 1, 1, 0, 4, },
19         { 2, 1, 2, 3, 1, 3, 4, }, { 1, 2, 1, 5, },
20         { 2, 0, 1, 5, 5, 3, 0, },
21         { 2, 0, 4, 2, 2, 4, 5, }, { 1, 3, 4, 5, }
22     };
23
24     // the 3-bit descriptor as natural number
25     int index = 0;

```

```

25 // calculating the descriptor
26 if( ascending )
27 {
28     if( vi[0][n] >= v[n] ) index |= 1;
29     if( vi[1][n] >= v[n] ) index |= 2;
30     if( vi[2][n] >= v[n] ) index |= 4;
31 }
32 else
33 {
34     if( vi[0][n] <= v[n] ) index |= 1;
35     if( vi[1][n] <= v[n] ) index |= 2;
36     if( vi[2][n] <= v[n] ) index |= 4;
37 }
38
39 int[] edge = sub_tri[index];
40 // key to calculate the subtriangles
41 int ntri = edge[0];
42 // number of subtriangles
43
44 bool found = 0;
45
46 // generate each subtriangle
47 for( int tri=0; tri<ntri; ++tri )
48 {
49     values nval[3];
50
51     for( int i=0; i<3; i++ )
52     {
53         int j = (tri*3)+i+1;
54         if( edge[j] >= 3 )
55             // no intersection, use old vertex
56             nval[i] = vi[edge[j]-3];
57         else
58         {
59             // field intersects triangle
60             // between v0 and v1
61             int v0 = (edge[j]) mod 3;
62             int v1 = (edge[j]+1) mod 3;
63
64             //relative intersection position t
65             double t = vi[v1][n] - vi[v0][n];
66             if( t != 0 )
67                 t = (v[n] - vi[v0][n]) / t;
68
69             // new value at this position
70             nval[i] = (1.0-t)*vi[v0] + t*vi[v1];
71         }
72     }
73
74     // recursively intersect subtriangles
75     // with next field n+1
76     if( hasCone( v, nval, ascending, n+1 ) )
77         found = true;
78 }
79
80 return found;
81 }

```


4.5.1 Parallel Computation of the Pareto Set

For each barycenter of a simplex, the *hasCone* code is run for each adjacent simplex of highest dimension. For example for the previous Figures 4.5 and 4.6, to identify the type of all points on an edge, excluding its two end points, the ascending and descending sets of the edge's barycenter have to be calculated in all adjacent triangles, usually two.

If, for example, the ascending set for x in any triangle is found to be non-empty, x cannot be a Pareto maximum. Otherwise, if all ascending sets in all adjacent triangles are empty, x has to be a Pareto maximum. As usual, this idea works analogously for descending sets and Pareto minima. The implementation guideline in Listing 4.2 therefore annotates each simplex, regardless of its dimension, with two boolean variables: *is_minimum* and *is_maximum*, with initial value *true*. When processing a d -simplex σ , the algorithm also processes each of its adjacent simplices σ' of dimension $d' < d$ by computing $H_{\sigma}^{+}(x)$ and $H_{\sigma}^{-}(x)$ for the barycenters x of each σ' using the clipping algorithm described above and sets, if necessary, their *is_minimum* and *is_maximum* booleans to *false*. When all d -simplices have been processed in that manner, the two booleans of each simplex of lower dimension are also set and classify it as either Pareto optimum, Pareto minimum, Pareto maximum, or as aregular simplex.

LISTING 4.2: Pseudocode to identify/classify Pareto extremal simplices in a simplicial complex.

```

void findPareto ()
2   {
4       forall( tri triangles ){
6           values val[3] = {
7               tri[0]->value,
8               tri[1]->value,
9               tri[2]->value,
10          };
12          // first, triangle itself
13          {
14              values vcen =
15                  barycenterOf( tri ).value;
16
17              if( hasCone( vcen, val, true, 0 ) )
18                  tri->is_maximum = false;
19
20              if( hasCone( vcen, val, false, 0 ) )
21                  tri->is_minimum = false;
22          }
23
24          // next, edges
25          for( int i=0; i<3; ++i )
26          {
27              int e0 = tri[(i+0) mod 3];
28              int e1 = tri[(i+1) mod 3];
29
30              // look up this edge
31              edge ed = findEdge(e0,e1);
32
33              values vcen =

```

```

34         barycenterOf( ed ).value
35
36         if( hasCone( vcen, val, true, 0 ) )
37             ed->is_maximum = false;
38
39         if( hasCone( vcen, val, false, 0 ) )
40             ed->is_minimum = false;
41     }
42
43     // check vertices
44     for( int i=0; i<3; ++i )
45     {
46         value vcen = tri[i]->v;
47
48         if( hasCone( vcen, val, true, 0 ) )
49             tri[i]->is_maximum = false;
50
51         if( hasCone( vcen, val, false, 0 ) )
52             tri[i]->is_minimum = false;
53     }
54 }

```

4.5.2 Interactive Computation of the Pareto Set

The definition of the Pareto set is based on information from all involved fields f_i . In detail, to decide if a point is Pareto extremal or regular, the intersections of the ascending set, $H_{i,\sigma}^+$, and the descending set, $H_{i,\sigma}^-$, need to be calculated for all $1 \leq i \leq n$. However, in this section it is shown that only a subset of fields is necessary to define these two sets, given piecewise linear fields over a simplicial complex. Hence, if a new field is added to f , the recalculation of the Pareto set requires only this subset and the added field, while the Pareto set based on the previous set of fields is already visualized as an interim result.

This observation provides two advantages to the Pareto set concept. It allows for a more efficient implementation of the *hasCone*-method than with the marching triangle algorithm for 2D domains, and thus allows a more effective data exploration. The effect of individual fields on the Pareto set visualization can be quickly explored by inserting or removing them to/from a already processed multifield. Also, the results for subgroups of the multifields can be efficiently saved and combined without a recalculation of the Pareto set based on the complete multifield.

To reduce the number of necessary recalculation steps upfront, the following two observations are to be considered: First, let $x \in \mathbb{M}$ be a Pareto optimum. For every $y \in U(x)$ in a sufficiently small, open neighborhood around x , there are two fields f_i, f_j such that $f_i(x) < f_i(y)$ and $f_j(x) > f_j(y)$. Hence, adding another field does not change x into a regular point. Therefore, only regular points need to be recomputed when a new field is added.

Second, let $x \in \sigma$ be a regular point such that the ascending set is non-empty, i.e. $H_\sigma^+(x) \neq \{x\}$. As stated in the previous section and in Chapter 1, the work focuses on 2D simplicial complexes with $f_i, 1 \leq i \leq n$, linear in σ . Thus, each $H_{i,\sigma}^+(x)$ is a half-plane restricted to σ and $H_\sigma^+(x)$ is a 2D-polygon with x as a corner. Obviously, each corner can only be adjacent to two edges and those are sections of either some coface of σ or the border of $H_{i,\sigma}^+(x)$, for some $1 \leq i \leq n$. This is true since $H_\sigma^+(x)$

is the finite intersection of all $H_{i,\sigma}^+(x)$ restricted to σ . For reference, the border of $H_{i,\sigma}^+(x)$ in σ that is also the border of $H_{\sigma}^+(x)$ is defined as

$$B_{i,\sigma}^+(x) := \{ y \in \sigma \mid (y - x) \cdot \nabla f_i(x) = 0 \wedge x \succeq y \}.$$

This definition implies that $B^{+i,\sigma}(x) = \{x\}$ iff $H_{i,\sigma}^+(x)$ does not "contribute" to the border. As a reminder, since all half-planes $H_{i,\sigma}^+(x)$ intersect at x , the 2D-polygon is actually a cone restricted to σ .

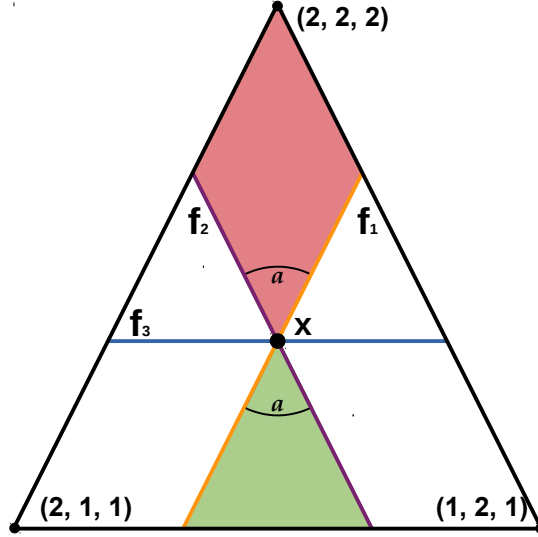


FIGURE 4.8: Example of a triangle with $f = (f_1, f_2, f_3)$.

Figure 4.8 provides an illustration of the two presented observations. Note that the third field f_3 does not contribute to the borders of the ascending or descending set which are colored in red and green, respectively. All three fields are linear inside the triangle σ and can therefore clearly be defined through the field values at the three vertices of σ . For the point $x \in \sigma^o$, the isolines $\{y \in \sigma \mid f_i(x) = f_i(y)\}$ for each field are drawn and labeled with their corresponding field f_i .

In higher dimensions, the ascending cone can be bordered by arbitrary many hyperplanes. For two dimensions, however, only two hyperplanes, given by two fields, are needed to define the ascending set. In this thesis, these indices with $B^{+i,\sigma}(x) \neq \{x\}$ are denoted as *defining indices* for σ .

Analogous considerations hold for the descending set. Due to the linearity of the field value distribution inside σ , the borders of both the ascending and descending set are point-symmetric with respect to x and defined by the same field indices. Hence, for each triangle, only two indices are needed to recompute both sets.

The thesis uses the above observations in the following algorithm. In a first iteration, considering at least two fields, the two indices i, j defining the borders of the ascending set are identified for each triangle. In the next incremental steps, it is sufficient to check if $H_{n+1,\sigma}^+(x)$ intersects the ascending set fulfilling either

$$B_{i,\sigma}^+(x) \cap H_{n+1,\sigma}^+(x) = \{x\} \text{ or } B_{j,\sigma}^+(x) \cap H_{n+1,\sigma}^+(x) = \{x\}.$$

If both conditions are true, f_i and f_j remain the most restrictive fields, i.e. i and j remain the defining indices. Hence, no additional field is needed to define the ascending set. Otherwise, if one of these conditions does not hold, i or j are replaced by another field as an additional defining index. If both terms are false, then $f_i(y) \geq f_i(x)$, $f_j(y) \geq f_j(x)$ and $f_{n+1}(y) < f_{n+1}(x)$ for all points y in the previously ascending set. Due to the point-symmetry of the sets, the reverse of these relations also holds for the descending set. Hence, x is surrounded by incomparable points and becomes Pareto optimal.

As a reminder, Lemma 4.5.1 shows that it is sufficient to apply this idea only to a single point inside each simplex, for example the centroid. Furthermore, let σ^* be the union of all cofaces of σ , i.e. edges and vertices. Based on the linearity of f_i it holds that

$$B_{i,\sigma}^+(x) \cap H_{n+1,\sigma}^+(x) = \{x\} \Leftrightarrow B_{i,\sigma^*}^+(x) \cap H_{n+1,\sigma^*}^+(x) = \{x\},$$

and

$$B_{j,\sigma}^+(x) \cap H_{n+1,\sigma}^+(x) = \{x\} \Leftrightarrow B_{j,\sigma^*}^+(x) \cap H_{n+1,\sigma^*}^+(x) = \{x\}.$$

Hence, for each triangle, the incremental step is reduced to the simple computation of line intersections along the edges. In detail, the intersection of the two points $B_{i,\sigma^*}^+(x)$ and $B_{j,\sigma^*}^+(x)$ with the lines H_{n+1,σ^*}^+ . For an efficient implementation H_{n+1,σ^*}^+ can be further separated into at most three subsections $H_{n+1,\sigma^*}^+ \cap e_k$, $k \in \{1, 2, 3\}$ with e_k the three edges of the triangle σ .

Figure 4.9 illustrates the idea of incrementally increasing the number of individual fields on a set of four artificial functions. It shows how the Pareto set grows by incrementally adding fields to the data set. In this example, the fields consists of Gaussian functions with minima in each corner and maxima that are located central at the top and bottom side of the domain. Note that the location of extrema is slightly shifted vertically and horizontally for each fields.

Note how the Pareto set in image (a) only consist of individual points colored in red or green. With only one field, the ascending and descending sets are half-spaces without any clipping and no incomparable points exist. For (b) - (d) the number of incomparable points increases which is reflected in the images by the amount of white areas in the triangles as well as the size of the Pareto set. Analysis of the images can, for example, consider that the increase of incomparable points going from image (c) to image (d) is less stronger than the increase between (a), (b), and (c). Hence, it is sufficient for additional investigations on how and where the functions differ or are most similar, to focus on the Pareto set for only the first three fields. Other important aspects that are revealed by an incremental analysis are, for example, that the fourth field mostly affects the upper left part of all connected components of the Pareto set in image (d) as those areas remain to regular regions in image (c), i.e. before field f_4 is added to the multifield.

The reverse order of images illustrates the idea of incremental removal of individual fields from the multifield. Other insertion or removal sequences would obviously result in other Pareto sets than shown in these images, excluding the final Pareto set, based on all four fields, which is invariant to these sequences.

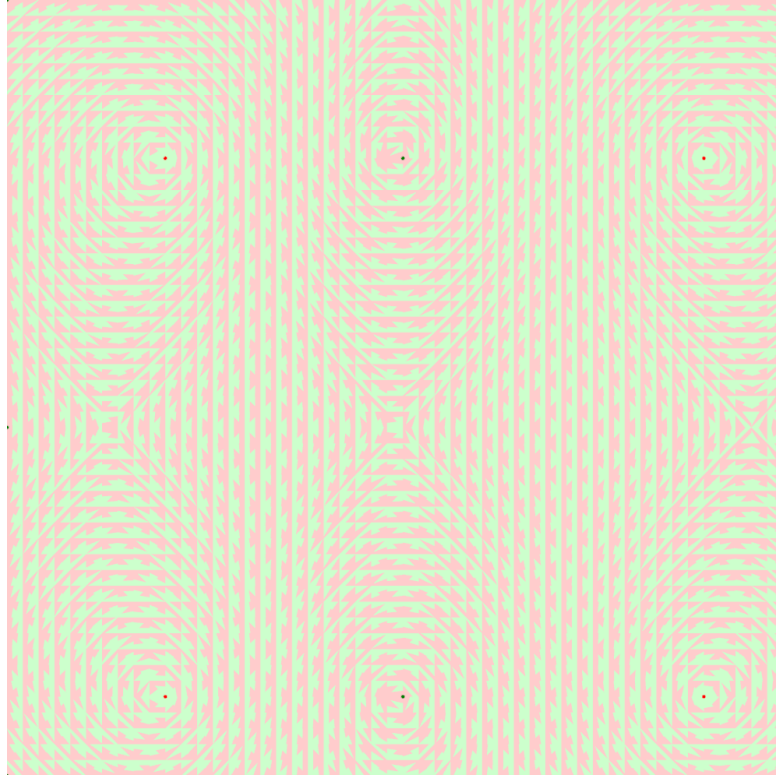
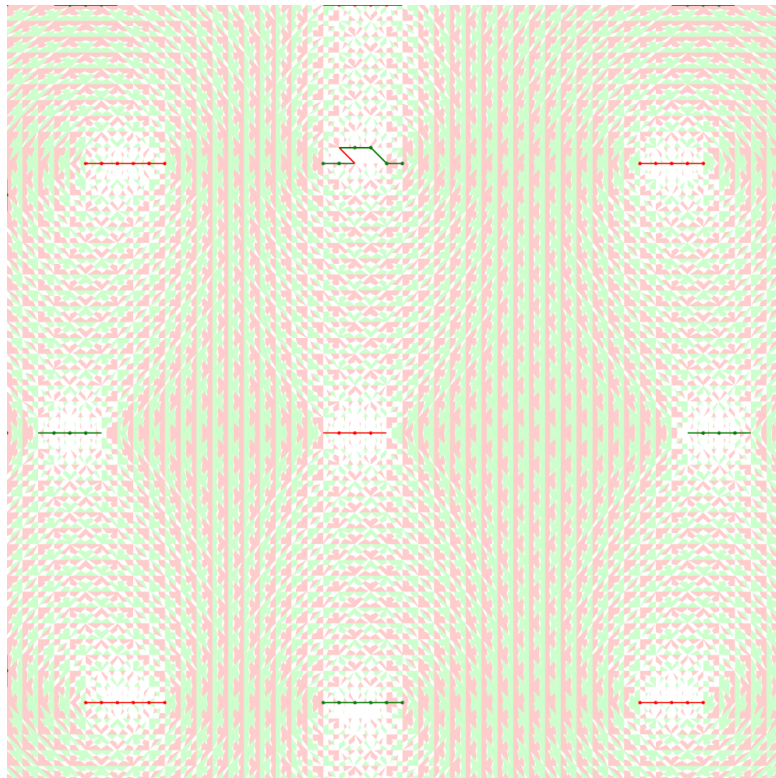
(a) f_1 (b) f_2

FIGURE 4.9: The Pareto set, colored in strong green, red and yellow according to their Pareto type. In each triangle the ascending and descending set for its barycenter are colored in less intensive colors according to the color scheme in previous Figures.

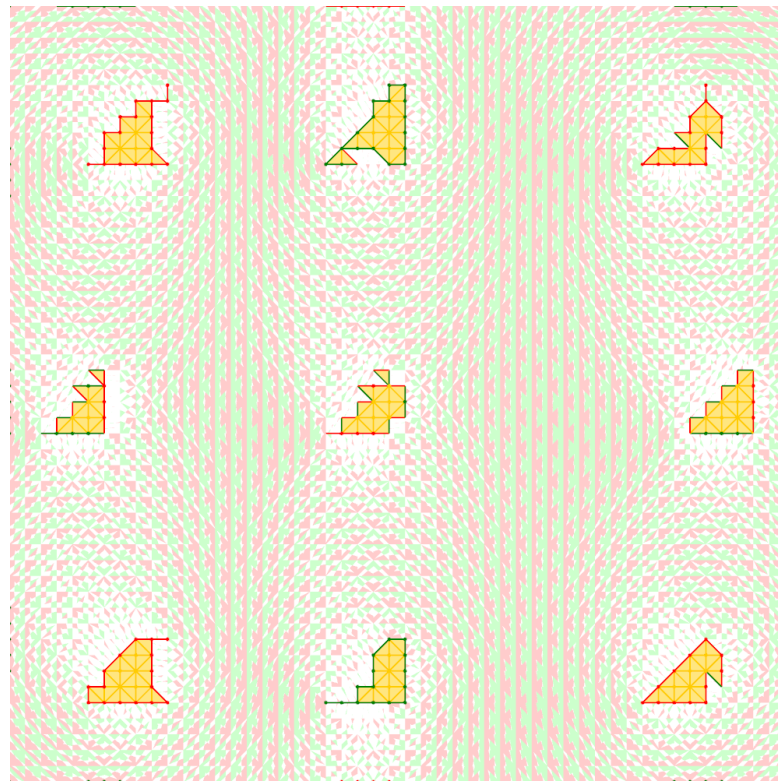
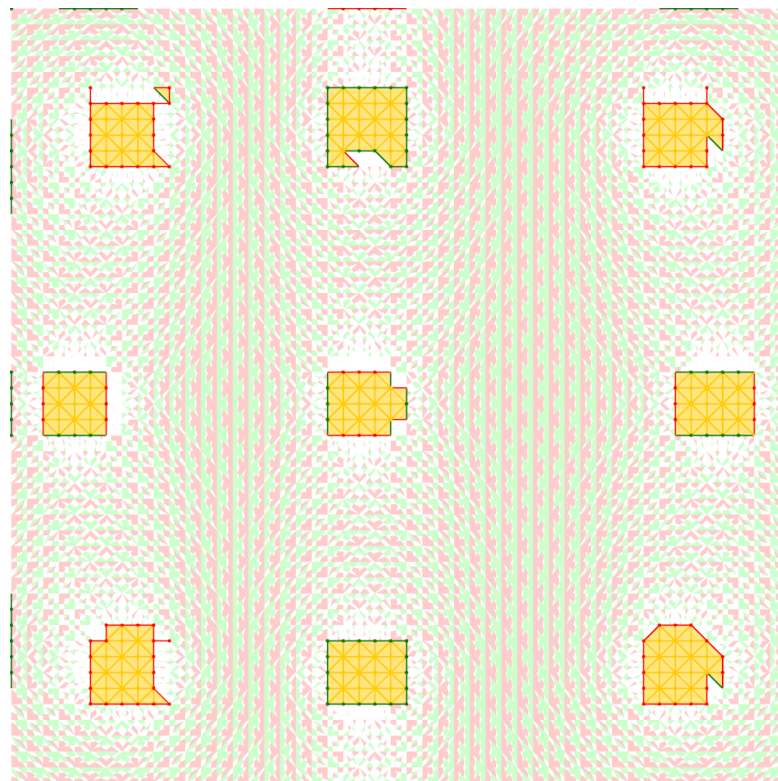
(c) f_3 (d) f_4

FIGURE 4.9: Pareto set and ascending and descending sets (cont.)

The implementation in 2D is outlined in the following pseudocode segment (Listing 4.3). The idea is that if the ascending cone for a point v inside a triangle exists, i.e. is non-empty, it intersects with at least one edge in form of an interval. The interval can be defined by its start and end point along the edge.

As done in the implementations in the previous subsections, let v be a vector saving the field values of a point at a triangle t and vi the matrix saving the field values of the three corners of t in its columns. The number of fields is given through $nFields$ and the flag *ascending* indicates if either the ascending or the descending cone is to be calculated.

LISTING 4.3: Iterative pseudocode to determine the ascending or descending set of a point in a triangle.

```

1  // for each edge of the triangle
   for( int j=0; j<3; j++ )
3  {
       // the edge from v0 to v1
       int v0=(j+0) mod 3;
       int v1=(j+1) mod 3;

       // initial B+_{i,\sigma*}
       // and B+_{i,\sigma*}
       double min = 0,max = 1;

       // for all fields
       for( int i=0; i<nFields; i++ )
       {
           // the relative position
           // where f_i(x) = f_i(v)
           double x = (v[i] - vi[v0][i]);
           x = x / (vi[v1][i] - vi[v0][i]);

           // H+_{i,\sigma*} it either from
           // [v0, x] or [x, v1]
           if( vi[v0][i] > vi[v1][i] )
           {
               // check if B+_{i,\sigma*}
               // intersects H+_{i,\sigma*}
               if( ascending )
               {
                   if( min<=x ){
                       min = x;
                   }
               }
               else
               {
                   if( max>=x ){
                       max = x;
                   }
               }
           }
           else
           {
               if( ascending )
               {
                   if( max>x ){
                       max = x;
                   }
               }
           }
       }
   }

```

```

47         else
48         {
49             if ( min<=x ){
50                 min = x;
51             }
52         }
53     }
54 }
55
56 if ( (max - min) > 0 )
57 {
58     // A (partial) cone is found
59     // intersecting the current edge.
60     return true;
61 }
62
63 // a cone does not exist
64 return false;

```

This implementation allows for the continuous addition of fields. An obvious extension to this implementation is to consider field deletion. However, removal of fields is more complex, since Pareto extrema can change their status and become a Pareto maximum, a Pareto minimum, or a regular point. While regular points obviously remain regular, their defining indices can change. If, for some triangle, the deleted field f_i is one of the defining indices, then the status of that triangle can only be determined if it is known (a) what the defining indices were before f_i was added, and (b) which fields were added after f_i . Note that all those field additions from (b) need to be recalculated sequentially, making the deletion of fields in the front of the adding sequence computationally expensive. The same recalculation can be done for Pareto extrema, assuming that before the addition of f_i the triangle previously contained regular points.

While it is in principle possible to store the sequence of defining indices to make field deletion interactive, i.e. allowing the removal of arbitrary fields, this is rather expensive due to the recalculation. Hence, it is reasonable to keep the sequence of defining indices for each triangle small to avoid the memory problems that motivated the incremental algorithm. It is also beneficial to only allow field deletion in reverse order of their addition to avoid recalculation, similar to an *undo*-option.

As a final remark towards interactive calculation, by saving the defining indices it is easy to combine the results of two separate multifields, given the same domain. Instead of recalculating the Pareto set for the union of the two sets of fields, it is sufficient to only consider the fields given by the two pairs of defining indices to calculate the new, collective Pareto set, i.e. to only use four fields in each triangle.

4.6 Discussion

The Pareto set concept has the advantage that the approach works for arbitrary numbers of individual fields without a restriction in terms of the domain dimension. This is especially advantageous in the context of visualization since the domains are usually two or at most three dimensional for such applications. Here, some existing topology-based approaches like the Jacobi set, are limited to only consider two or three fields. However, it is often useful to consider more than just two or three fields,

for example to better compensate or analyze uncertainty.

Furthermore, the results can be visualized directly within the domain and thus combined with other maps, like selected individual fields or geographical information of the domain, for example in the context of atmospheric research. The visualization of the Pareto set and the complementing spatial information can be done, for example, by two different media like color and height.

The fields are only compared according to their topology and gradients. Hence, the Pareto set is invariant to scaling factors or measuring units of the individual fields. However, the results remain sensitive to field inversion, i.e. the Pareto set changes if instead of a field f_i , its inversion $-f_i$ is considered. This detail makes Pareto sets very suitable to maximal and minimal behavior among multiple different scalar fields, as long as the fields are semantically related.

4.6.1 Analytical Examples and Visualization Ideas

To demonstrate the definition presented in this chapter and to indicate its relevance for analysis purposes, the following analytical examples are presented. A general concept, used in this thesis, is to color Pareto extrema based on their type. In detail, Pareto maxima are colored green, Pareto minima are colored red, and Pareto optima are colored yellow. This can be done either for each simplex separately or for each component in $\mathbb{E}(f)$. For 2-simplicial complexes, the local ascending sets, colored red, and local descending sets, colored in green, for the barycenter of each triangle can be used to visualize areas with regular points. For 3-simplicial complexes the visualization of the ascending and descending sets would obscure the Pareto set. Hence, for 3D domains regular points are either omitted from the visualization or indirectly presented with the reachability graph.

The created artificial analytical data set are sums of Gaussian bell functions as presented in the corresponding section in Chapter 3 and approximated to piecewise linear fields over a simplicial complex.

The first data set in $2D$, depicted in 4.10(e), is based on four fields, each with two positive and two negative peaks and perturbed centers. Since only boundaries, i.e. sequences of lines and points, can be Pareto minimal or maximal, the enclosed Pareto optima can alternatively colored using the following code:

- light red for regions with mostly minimal Pareto boundary,
- light green for regions with mostly maximal Pareto boundary,
- and yellow for all other regions.

Note that this code approximates the definition of Pareto extremal nodes in the reachability graph based on in- and outgoing edges. It is only an approximation and equality cannot be guaranteed, however this visualization does not require the calculation of the reachability graph and is therefore much faster.

Instead of the local ascending and descending sets, in Image 4.10(e), the visualization is based on line-integral convolution (LIC) [27] of the angle bisection of $H_\sigma^+(x)$ at each position x to depict a very rough approximation to the overall fields' gradient behavior. In detail, the angle bisections of $H_\sigma^+(x)$ for regular points create a

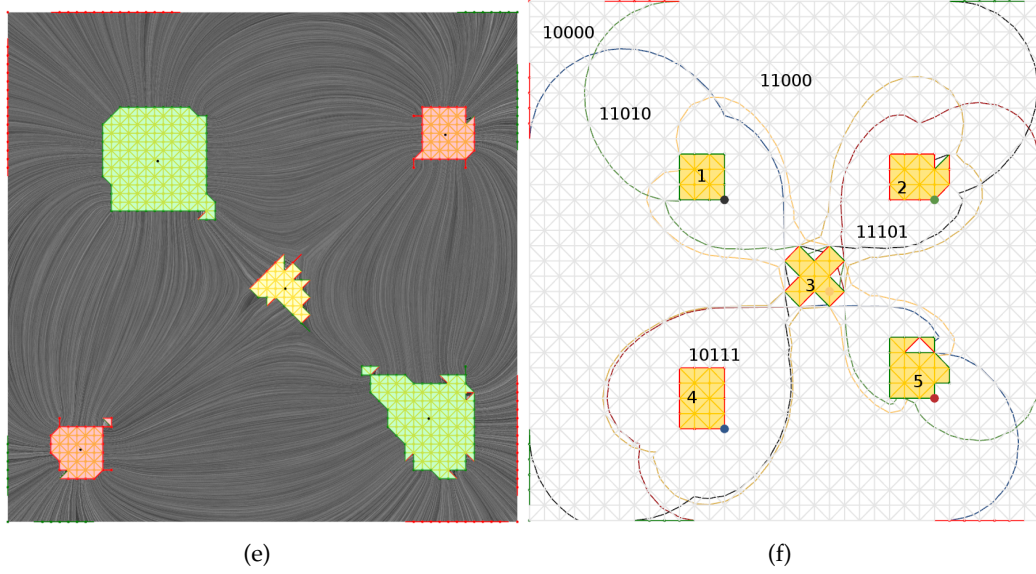


FIGURE 4.10: Analytical examples. (a) Pareto extrema of four 2D fields, each with two positive and two negative peaks, and line-integral convolution depicting approximated gradient directions in non-Pareto regions [27]. (b) Pareto extrema along with the global ascending and descending sets for the five Pareto optimal regions. For illustration, some global ascending/descending sets intersections have been annotated with a bit string indicating the relation to the Pareto extremal regions.

(piecewise constant) vector field. Given a randomized gray level image N , i.e. each pixel has a random gray scale value, of the same size as the domain \mathbb{S} , each pixel x is colored with the following map m :

$$m(x) := \int_{-L}^L k(s) N(p_x(s)) ds$$

with p_x the path going through x , i.e. $p_x(0) = x$, following the vector field, and $2L$ a considered the path length. The kernel k defines the weights along the path p . Common kernels found in mathematical literature [170, 11] for $|u| \leq 1$ are for example:

Uniform (box) $k(s) = 1/2$,

Triangular $k(s) = 1 - |s|$,

Epanechnikov $k(s) = \frac{3}{4}(1 - u^2)$,

Quartic (biweight) $k(s) = \frac{15}{16}(1 - u^2)^2$, or

Gaussian $k(s) = (2\pi)^{-\frac{1}{2}} \cdot e^{-\frac{1}{2}u^2}$.

For the corresponding image in Figure 4.10, the simple box kernel is used. The general idea of LIC is that close points along the same path receive a similar value since their are integrated over almost the same pixels. Otherwise, close points that are not on the same path, i.e. on a parallel path, are integrated over different image pixels, thus receive different values. Hence, LIC creates an image consisting of lines that run

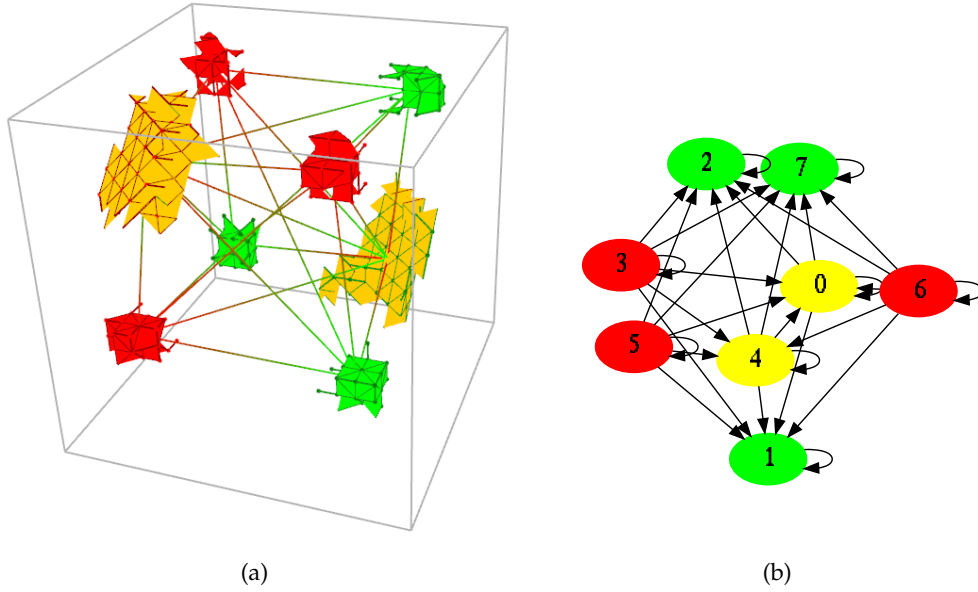


FIGURE 4.11: Analytical examples. (a) Pareto extrema of eight 3D fields, each based on 3D Gaussian bells with three positive and three negative peaks. The reachability graph is placed in the domain. The nodes are placed at the gravity center of their corresponding elements of $\mathbb{E}(f)$. (b) A graph layout that shows the reachability graph of the multifield in (c). Generated by C.Heine with graphviz's fdp [63].

parallel to a given gradient field. This concept is easy to implement and fast to compute since the calculations are local and can be limited with $2L$. It is a basic approach to visualize vector fields.

The image shows five Pareto optimal regions and some Pareto minimal and Pareto maximal regions at the boundary. Note that the Pareto optimal regions have uneven boundaries and that the boundaries are not exclusively minimal or maximal although the used Gaussian field would suggest the opposite, at least around the location of minima and maxima of the individual fields. A possible reason for those artifacts is the use of the piecewise linear approximations.

Points that are Pareto optimal are incomparable to their neighbors. Therefore, a Pareto optimal region shows the lack of a clear *consensus* between the individual scalar fields. Note that the size of a Pareto optimal region is a first visual indicator on how close the fields are to a consensus. In detail, the image shows that the fields have less common behavior around the two red Pareto minimal regions than around the green Pareto maximal regions.

In Figure 4.10(f), the Pareto extremal regions are shown for another 2D data set, but this time also the global ascending and descending sets are shown for each Pareto optimal region. These are depicted by their colored boundary. Note that each Pareto optimal region contains a large circular marker with a color equal to the boundary color of the corresponding global ascending and descending set. For example, the blue boundary corresponds to the Pareto optimal region marked with "4" in the lower left. Similar to Morse-Smale cells, an equivalence class can now be defined

for each point based on the set of ascending and descending sets that contain them.

While these visualization strategies work for 2D, in 3D examples it is only feasible to show the boundaries of Pareto extremal regions as solid surfaces, as demonstrated in Figure 4.11(a). This image shows another artificial data set that uses eight fields each with three positive and three negative peaks, resulting in three Pareto maximal and minimal regions, respectively. In the visualization, also two saddle-like Pareto extrema emerge, thus in total eight connected components in the Pareto set. An approach that works for any dimension is to show the reachability graph of ascending and descending sets, as can be seen in Figure 4.11(b) for the same 3D data set. The components in Figure 4.11(a) are classified and colored based on the number of ingoing and outgoing edges of their corresponding nodes in the reachability graph, as defined previously in this chapter. Again, loops exist in the graph due to Pareto minimal points reaching Pareto maximal points of the same connected component and are based, most likely, on the inaccuracy of the triangulation.

4.6.2 Limitations

While the definitions work for continuous fields and derived simplicial complexes, they have a number of limitations with regard to the data and the visualization.

Note that in the single-field case, where only one field $f = (f_1)$ is considered, points cannot be incomparable. Therefore, every point y in a neighborhood of a Pareto maximum x has to be dominated by x , i.e. $x \succeq y \Rightarrow f_1(x) \geq f_1(y)$, which is equivalent to x being a maximum for f_1 in terms of single-field topology. This equivalence also works for Pareto minima and single-field minima. However, the definition of ascending and descending sets does not consider the number of connected components in those sets. Thus, the definition does not distinguish between regular points and those corresponding to so-called saddle points from single-field topology. Nevertheless, the above illustrations (e.g. Figure 4.10) demonstrate that single-field saddles, closely positioned in the domain but in different fields, can result in saddle-like structures.

Furthermore, data sets with a large number of small, localized structures and flat regions contained in individual fields pose limitations to the method. Assume a local maximum x for some field f_i , i.e. for a neighborhood $U(x)$ around x , it holds that $\forall_{y \in U(x)} f_i(y) < f_i(x)$. Thus, x cannot be a regular point since this requires at least on neighbor $y \in U(x)$ with $f_j(y) \geq f_j(x)$ for all $1 \leq j \leq n$, including $j = i$. Obviously, this holds for minima alike. As a result, every local feature of any single field creates a feature in the Pareto set visualization or, generally speaking noise from the individual fields accumulates into the multifield.

On the other side, a non-critical point x with respect to a single field f_i can be a Pareto extremal point. This is especially true if there is another field f_j with $f_i = -f_j$ for some neighborhood around x . Hence, all points inside that neighborhood are Pareto extrema and create a large connected component in the Pareto set. Hence, the Pareto set approach is more suitable for multivariate data, i.e. variables with a similar semantic. This is the case for example if all variables are criteria for the same phenomenon or for ensemble data sets.

Obviously, other types of multifields without a common domain, such as point data or a set of simulation results with different spatial resolutions, have to be pre-processed and retriangulated to create the same domain for all fields.

Another limitation is the restriction to smooth Morse function. It is necessary to define the continuous ascending and descending paths, enforces that every vertex has only one set of scalar values, and also avoids problems with flat areas. In a flat areas all fields are constant such that for any point x in that area $\forall_{y \in U(x)} f(x) = f(y)$. Hence $x \succeq y$ and also $y \succeq x$ for all points $y \in U(x)$, making x theoretically both Pareto maximum and Pareto minimum but not a Pareto optimum. This is problematic since it contradicts the case for non-flat regions where every Pareto optimum is also Pareto maximum and Pareto minimum. Fixing the definition to allow for flat regions is difficult however, since it is impossible to identify, based on local neighborhood information, if a flat area corresponds to a valley or a plateau in the data and thus should be considered minimal or maximal, respectively.

4.6.3 Runtime

It is easy to see that in the 2D case with simplicial complex representation of the data, the runtime is in $O(3 \cdot n \cdot 7 \cdot t)$ with n the number of fields and t the number of triangles. For each simplex (triangle, edge, or vertex) the ascending and descending set in all adjacent triangles has to be calculated. Only for one point per simplex the Pareto type needs to be computed to determine the type of all points in that simplex. With the implementation of *hasCone* in 2D, for each triangle all three edges have to be considered, resulting in a factor 3 for the runtime.

For higher dimensions however, only the merging triangle algorithm is at hand. In this case, the runtime is much longer because, in general, a clipped simplex is no simplex, but can always be represented by the union of at most d simplices which are stored as a list of simplices instead of a polyhedron. As each additional half-space intersection potentially increases the number of simplices in this list, the worst-case runtime is in $O(d^k)$.

Note that this runtime only holds for the calculation of the Pareto set. The reachability graph needs even more time to create, as the number of global ascending sets is extremely high. The implementation for 2D domains used for this chapter creates the global ascending set for each connected component in the Pareto set separately. Each edge at the component border, with vertices a and b , receives a value vector v with $v_i = \min\{f_i(a), f_i(b)\}$ for $1 \leq i \leq n$. Since the fields f_i are piecewise linear, v is sufficient to describe the front of the ascending set. The value vector is passed through the adjacent triangle to the neighboring edges to push the front of the currently calculated ascending set forwards. At each edge, with vertices a' and b' , the value vector is updated. W.l.o.g. let $f_i(a') \leq f_i(b')$, then

$$\begin{aligned} v_i^{\text{new}} &= v_i & \text{if } f_i(a') \leq v_i \leq f_i(b'), \\ v_i^{\text{new}} &= f_i(a') & \text{if } v_i \leq f_i(a'), \end{aligned}$$

and otherwise the front does not reach that edge and is not further propagated. If v is updated however, the ascending front is pushed to the next triangle. Each edge is visited by the front of the ascending set in the worst case by each of its neighboring edges, at least four times, the running time of the front propagation is in $\Omega(E \cdot E' \cdot 4)$ with E the number of all edges in the simplicial complex and E' the number of all

Pareto minimal edges. The running time for the implementation for 3D domains is analogously $\Omega(T \cdot T' \cdot 4)$ with T and T' the number of all triangles and all Pareto minimal triangles, respectively.

Note that the ascending front can visit each edge repeatedly with an updated value vector, thus creating an even longer running time, even if piecewise linear fields are considered and an assumption is possible that many paths run parallel and thus can be summarized. It is however possible to adapt ideas from Morse-Smale complexes and their calculation, since both concepts are slightly similar. However, as shown in numerous works for Morse-Smale complexes, e.g. [77, 175, 148, 75, 149] their calculation is rather complex and time consuming, too. Nevertheless, options are discussed in this dissertation to reduce this run time problem in Chapter 6.3. Also, as stated above, details like edge placement and the visualization as a multi-graph is still an open discussion.

Chapter 5

Simplification of Pareto Sets

As stated in the previous chapters, Pareto sets, and multifield visualizations in general, are prone to cluttering due to many small, local structures in the data. All local structures in the individual fields occlude the large multifield structure, and can hence be considered as *noise*.

The removal of those local structures is often done in two steps. First, a decision function is applied to distinguish noise from potentially important information. Then, the multifield data set is modified to remove the marked structures either implicit in the visualization or explicit in the individual fields itself.

The presented multifield simplification methods described in the following sections are results from previous work, found in previous work [86].

5.1 Using Existing Single-Field Approaches

A straight-forward approach removes unwanted local structures in the individual fields before the multifield analysis is applied. This also avoids computational overhead in the latter.

For the single-field case, several approaches exist to either smooth or simplify an individual field [138, 41, 50, 78, 72, 77, 162, 163]. To remain in the context of topology-based methods, consider a Reeb graph or contour tree of a given data set (see Section 3.2.2). Implicit simplification of the data is done through the simplification of the graph structure, i.e. by merging two adjacent nodes into one [5]. As a reminder, each node in the graph represents a critical point in the single-field topology (minimum, maximum or saddle point). Each edge in the graph can be associated with the value difference between its two corresponding nodes. This difference, called *persistence*, can be considered the amount of *work* an explicit simplification of that edge would cost and can be used for a decision function.

```
sort all edges based on persistence
starting from the lowest edge:
    merge current edge
    repeat until no edge with persistence < minP exists.
```

Note that a merge of two points might also change the persistence of their adjacent edges. For example, consider a path a, b, c, d in a larger Reeb graph. The distance $d(a, d)$ is equal to $d(a, b) + d(b, c) + d(c, d)$ and should stay the same after a merge of b and c into a single node c' , since a simplification should only remove

local structures. Hence, the new distances $d'(a, c')$ and $d'(c', d)$ have to be modified. For example, let $d'(a, c') = d(a, b)$ and $d'(c', d) = d(b, c) + d(c, d)$. This is valid since only points with the lowest persistence are merged such that no negative values are created.

Note that it is possible that the new node c' is regular in that Reeb graph, i.e. only has one ingoing and one outgoing edge. Therefore c' can be removed from the Reeb graph completely, by connecting a and d directly with the new distance $d'(a, d) = d'(a, c') + d(c', d)$.

The implicit changes of the persistences at the edges also reflect the explicit changes of the scalar values in the data. Two common approaches to do these data changes are called *carving* and *flooding*. The latter is illustrated in Figure 5.1, in which a small feature in a 1D function, created by a small peak x and a sink y , is removed. The two critical points, corresponding to the peak and sink are marked by red and green dotted lines, respectively. Flooding means that all points adjacent to y with a scalar value between $f(y)$ and $f(x)$ are changed to $f(x) + \epsilon$ with $\epsilon > 0$ inducing a small descend towards x . Carbing means that all points adjacent to x with scalar values between $f(y)$ and $f(x)$ and changed to $f(y) - \epsilon$ with $\epsilon > 0$ inducing a small ascend towards y . For further explanation and for information on an efficient implementation of a simplification algorithm using Reeb graphs, see work by Tierny and Pascucci [162].

In contrast to guided simplification methods based on global structures, *data smoothing* changes the data only based on local information. The method used in later sections is derived from Gaussian blur [125], an image processing tool to reduce image noise and reduce detail. The function value at each point is set to a weighted average of the value of that point's neighbors with the Gaussian function as weights.

As a reminder, the Gaussian function in two dimensions, where x and y is the distances from the origin in the x - and y -axis, respectively, and σ the standard deviation:

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}.$$

The original point value receives the largest weight with a distances of zero, while the neighboring points receive smaller and smaller weights as their distance increases. Since the weights decrease exponentially with the distance, this smoothing technique preserves boundaries and sharp edges in the data. Furthermore, the weights converge towards zero exponentially fast so that the implementation uses a distance variable after which all weights are approximated by zero. Hence, only points inside a set neighborhood need to be considered for the weighted average.

5.2 Simplification based on Reachability Graphs (RGS)

While Chiang et al. [38] apply their contour tree-based smoothing technique to single-field scenarios, an adapted approach for multifields is used in this thesis. For this approach, the global structure as represented by the reachability graph (see Section 4.4)

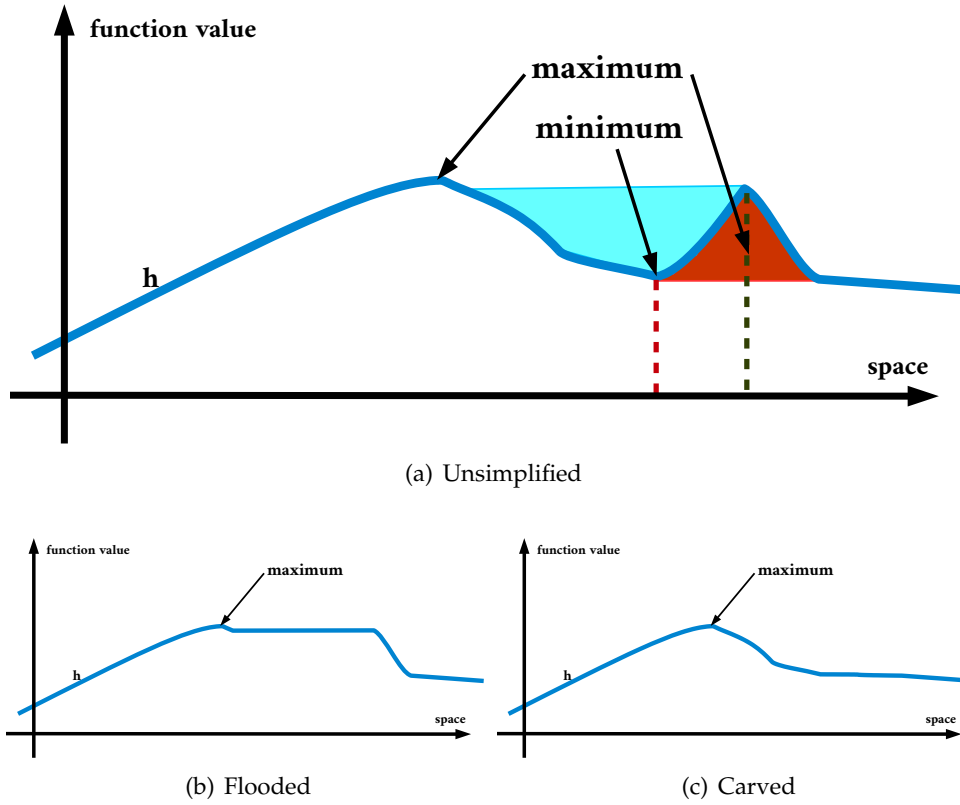


FIGURE 5.1: Illustration of two single-field simplification methods. (a) shows the unsimplified function and indicates the regions that are to be flooded and carved as blue and red areas, respectively. (b) shows the simplified function, if the flooding concept is applied, while (c) depicts the simplified functions after the peak is carved away.

is used to provide an abstract view on the domain-wide connections between connected components of Pareto extrema, i.e. the elements of $\mathbb{E}(f)$. Simplification operations for this graph and how those operations are translated to actual modifications to the data are discussed in this section. Additionally, following the two steps simplification idea indicated in the introduction of this chapter, a suitable criterion to find a good sequence of graph simplification operations are defined.

Given the reachability graph $RG(f)$, there are two types of possible operations that yield simplification of the graph and are therefore suitable to simplify the structure of $\mathbb{E}(f)$. All other operations to a graph structure can be done through a sequence of these two or their reverse actions.

Merge two adjacent nodes into a new one, which inherits the old nodes' edges and neighbors.

Purge a node and remove it along with every adjacent edge from the graph.

In detail, let $RG(f) = (V, E)$ be a directed graph with nodes V and edges $E \subseteq V \times V$. To merge $a, b \in V$ into some $c \notin V$, the new graph $RG' = (V', E')$ with $V' = \{c\} \cup V \setminus \{a, b\}$ and $E' = E|_{a \leftarrow c, b \leftarrow c} \setminus \{(a, b), (b, a)\}$ is created. The equation

$E|_{a \leftarrow c, b \leftarrow c} \setminus \{(a, b), (b, a)\}$ means that every appearance of a or b in the set of edges E is replaced with c and every loop that might get created through this renaming is removed.

To purge $a \in V$, the new graph $RG' = (V', E')$ with $V' = V \setminus \{a\}$ and $E' = \{(x, y) \in E \mid x, y \neq a\}$ is created. However, removal of any node with more than one adjacent edge can result in a disconnected graph, which implies disconnected data. However, since every reachability graph is connected, a purge operation must be avoided during simplification. The only allowed purge is the removal of a leaf, which is topologically equivalent to a merge operation of the leaf and its adjacent node. Hence, for the graph simplification approach only merging of nodes needs to be discussed.

5.2.1 Interpretation

A **merge** operation in RG can have different interpretations in terms of how the function f changes such that the new reachability graph RG' is isomorphic to the original $RG(f)$ after the merge operation. For a contour tree representing a single field, there are several ways as is, for example, presented by Tierny and Pascucci [162]. To provide a broad overview, three approaches to change f into f' which corresponds to a merge of the reachability graph $RG(f)$ into $RG(f')$ are described. Note that more efficient methods might be possible but remain open discussion.

Given an edge (R_1, R_2) connecting two connected components of $\mathbb{P}(f)$ R_1 and R_2 , either R_1 or R_2 can be removed, or these components are merged through a connecting region of Pareto extrema. Figure 5.2 illustrates the changes to f that are equivalent to a merge in the corresponding reachability graph $RG(f)$. Assume that R_1 and R_2 are elements in $\mathbb{E}(f)$ with the corresponding nodes R_1 and R_2 and an edge (R_1, R_2) . The hatched red and green areas are sections of the ascending and descending set of R_1 , the unhatched are those of R_2 . The blue encircled region describes the equivalence class of a regular point $H^+(R_1) \cap H^-(R_2)$. It contains all ascending paths that start from a point in R_1 and end somewhere in R_2 . To merge the nodes R_1 and R_2 , they can be connected by Pareto extrema or one of them can be replaced by regular points. In either case, it is not allowed to create any more connected components in $\mathbb{E}(f)$ or loops in the reachability graph. Otherwise, the changes to f do not correspond to the merge operation in the reachability graph.

Figure 5.2(b) illustrates the removal of R_2 . It shows the ascending and descending sets of R_1 . Note how all paths from or to R_2 are now re-connected to R_1 without creating circular paths or new Pareto extrema.

The functions are modified in such a way that all function values of the points in R_2 are iteratively moved towards R_1 . Figure 5.2(c) illustrates the idea behind this *pushing* analogy on a one-dimensional example with two functions (orange and blue) and two Pareto extrema with the same color scheme as in previously shown images and as described in Section 4.6. The function values in the neighborhood of moved points are interpolated to create the topological behavior as illustrated in Figure 5.2(b).

The second option, removal of R_1 to merge (R_1, R_2) in the reachability graph, is symmetric to this illustration. However, these two options are not able to handle

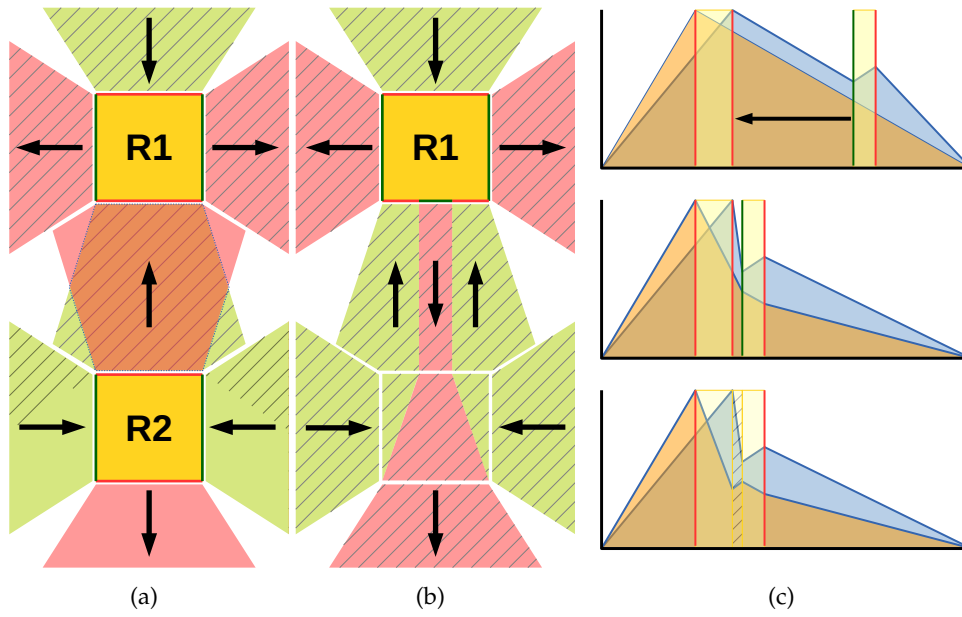


FIGURE 5.2: Illustration of the functional change if an edge corresponding to a connection between $R1$ and $R2$ is merged. (a) and (b) show the result of the changes to the ascending and descending sets of $R1$ while (c) provides implementational ideas through a 1D example.

loops, i.e. when $R1 = R2$. Because the removal results in regular points that do not reach Pareto maxima or Pareto minima. This is a contradiction to the above observation.

Hence, for the third option, i.e. connecting $R1$ and $R2$ through additional Pareto extrema, the functions inside the blue circled area in Figure 5.2(a) are modified. All ascending paths from any point of $R1$ to some point of $R2$ are contained in this area. While a single line of Pareto extrema from some point in $R1$ to another in $R2$ also connects these components, thus forcing the change of all points in the encircled regions since otherwise loops from this connected component $R1/R2$ to itself are created. This is, however, not topologically equivalent to the merge of $(R1, R2)$. Hence, all points inside the blue circled area are changed into Pareto extrema through a modification of the multifield f .

As an example, consider the image from Chapter 4, also shown in the Figure 5.3, with four vertices connected in a circle as a simple and compact illustration. The vertices v_1 reach v_3 via a path through v_2 . A point along this path is chosen, here v_2 , and change at least two of its function values, here f_0 and f_1 . The first value is decreased below $\min\{f_0(v_1), f_0(v_3)\}$ and the second value is increase above $\max\{f_1(v_1), f_1(v_3)\}$, e.g. $f'(v_2) = (0, 4)$. The function values for v_1 and v_3 remain unchanged while all other points along the path are linearly interpolated. Note that all points along the path become Pareto extremal.

In summary, two key elements for the simplification are illustrated in Figure 5.2(a). First, given the area that contains all ascending paths from one element of $\mathbb{E}(f)$ to

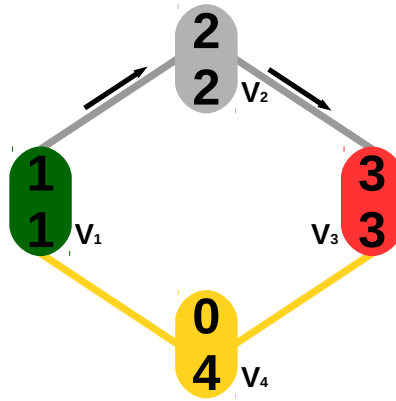


FIGURE 5.3: A simple example with two sets of scalar values over four vertices and four edges. Vertex v_1 is a Pareto minimum, colored green, v_3 is Pareto maximal, colored red, and the yellow colored vertex v_4 and adjacent edges are Pareto optimal. v_1, v_3 , and v_4 and all points along the mentioned edges build on connected component S , i.e. one element in $\mathbb{E}(f)$.

the other – the blue circled area in Figure 5.2(a) – all function values at the boundary of that area and outside are not changed. Second, at least two functions are modified in such a way that their gradient vectors are inverted to each other for all points inside that area.

5.2.2 Operation Sequence

The last step towards a viable simplification method is to determine a suitable order of merge operations such that simplification addresses small-scale regions (e.g. resulting from noise) with preference.

The previous paragraph outlines the merge of two nodes (a, b) either through removal of $R2$, removal of $R1$, or through an additional connection between $R1$ and $R2$. The first option modifies only a very small part of $H^+(R1) \cap H^-(R2)$ and all points of $R2$, the second option only some points in $H^+(R1) \cap H^-(R2)$ and all points of $R1$ are changed, and for an additional connection between $R1$ and $R2$ it is required to change all points in $H^+(R1) \cap H^-(R2)$.

To estimate the amount of change to f required to transform a regular point x into a Pareto optimum, the dominated and/or dominating points with respect to x restricted to a small neighborhood of x are measured, i.e. $H_\sigma^+(x) \cup H_\sigma^-(x)$ as defined in Section 4.1, σ being the simplex containing x . If those points become incomparable to x by small changes to f , x becomes Pareto optimal.

As a reminder, since the multifield f is assumed to be linear in σ , $H_\sigma^+(x)$ is an intersection of n convex half-spaces, each with a hyperplane as border defined by $\{y \in \sigma \mid f_i(x) = f_i(y)\}$ for $1 \leq i \leq n$ and σ .

Definition 5.2.1. Let x not be Pareto maximal. $H_\sigma^+(x) \setminus \{x\} \neq \emptyset$ is a convex polyhedron and the cone at x has an opening angle which reflects the local quantity of dominated points invariant to the size of σ and the actual scalar values from f . In this thesis, this angle is

denoted as $\alpha_\sigma^+(x) \in [0, \pi]$. If x lies on a face of some simplex, σ becomes a set of simplices to which x is adjacent and α^+ is calculated as the sum of opening angle in those simplices,

$$\alpha^+(x) := \sum_{\sigma \in T} \alpha_\sigma^+(x),$$

with T the set of all d -simplices containing x .

Due to the piecewise-linearity of the individual fields f_i , the value α^+ is constant for each simplex. The piecewise-linearity also implies that inside d -simplices, the ascending and descending sets are point symmetric. Hence the opening angle for Pareto maxima, for which the ascending set is empty, is defined through the descending cone, i.e. $\alpha_\sigma^-(x)$ or $\alpha_\sigma^-(x)$. If x is Pareto optimal, both ascending and descending sets are empty. It is therefore reasonable to define $\alpha_\sigma^+(x) := 0$ for all Pareto optima x , although the opening angle does technically not exist.

In this thesis, the amount of required alterations to f to transform all points $R = H^+(R_1) \cap H^-(R_2)$ into Pareto extrema is estimated using $\int_{y \in R} \alpha_{\sigma_y}^+ + \alpha_{\sigma_y}^- dA_y$. However, note that the ascending and descending set for x are point symmetric at x . Therefore $\alpha_\sigma^+(x) = \alpha_\sigma^-(x)$ holds. Hence, $\int_{y \in R} \alpha_{\sigma_y}^+ dA_y$ equivalently reflects the alteration to f except for a scalar factor.

To change a Pareto optimum x into a regular point, an optimization problem must be solved for each point, which is only practical for small problem instances. An obvious upper bound with the same approach as α_σ^+ is to measure all incomparable points with respect to x restricted to a small neighborhood around x . For Pareto optima, this is obviously the complete neighborhood, therefore an upper bound.

Definition 5.2.2. *The reachability graph is augmented with the following weights such that it becomes a weighted reachability graph $RG_w(f) = (V, E, w)$. For each edge $(s, t) \in E$, corresponding to components $S, T \in \mathbb{E}(f)$, $w(s, t) = \int_{y \in R} \alpha_{\sigma_y}^+ dA_y + \min\{|S|, |T|\}$ is defined with $R \subseteq H^+(S) \cap H^-(T)$ as described above.*

With this definition, $w(R_1, R_2)$ is an estimate of the work needed to remove the component R_2 in the previous image. Furthermore, as the node R_2 is removed, all incoming and outgoing edges of R_2 become incoming and outgoing edges of R_1 . To adjust the weights without a recalculation of the ascending and descending sets, $w(R_1, R_2)$ is added to every relocated edge weight and sum the weights of multiple edges, if those appear. The weights measure how much the functions at a point have to be changed with respect to the halfplanes to turn a regular point into a Pareto extremum. Note that these weights remain conservative estimations since in the piecewise linear setting the change applied to one point also changes all other points of the same simplex.

A greedy algorithm is applied to find a sequence of edge merges until a given threshold is reached. In particular, the algorithm chooses the edge with minimal weight, merges the two adjacent nodes as illustrated before by merging one of the connected components into the other. Then the algorithm recomputes the remaining weights and repeats the procedure until no edge with a weight smaller than the given threshold can be found.

As noted in Section 4.6 through a set of analytic examples, errors in the triangulation might create errors, i.e. wrongly classified Pareto extrema and loops in the

reachability graph. However, these errors are small and local, i.e. they usually affect only single simplices σ and the opening angles $\alpha_\sigma^+(x)$ are rather small. Hence, loops that are created through these triangulation errors have a small weight and are the first ones to be removed by the greedy algorithm, via a merge of the adjacent node with itself.

5.2.3 Implementation

The current implementation of computing the angle is limited to two dimensional domains, specifically simplicial complexes (triangulated grid) with scalar values given at the vertices. The same algorithm as in Chapter 4 is used to identify the set of Pareto extrema $\mathbb{E}(f)$.

The idea is to compute the ascending and descending set for each component using the following algorithm. It begins at the Pareto minima, respectively Pareto maxima, of the components and propagates a front along the ascending, respectively descending, paths, extending it through individual triangles.

If the front encounters a Pareto extremum and a corresponding edge that is not already in the reachability graph, such an edge is added. This is done until every section of the front has stopped at a Pareto extremum. In each triangle, the corresponding section of the global ascending or descending, set is saved such that in a second step, the weights for each edge can be calculated in each triangle individually and then be accumulated. Note that the sections of the ascending or descending sets, respectively, in the triangles are defined by at most six points, two along each edge of the triangle; in other words, in the worst case, each triangle is visited by the front three times. This makes the calculation of intersections straightforward.

Let $c := |\mathbb{E}(f)|$ be the number of connected components. The reachability graph $RG(f)$ has at most c^2 edges, the calculation of the ascending and descending sets and their intersection can be broken down to n linear problems in each of the m triangles such that a worst-case running time in $O(c^2 \cdot m \cdot n)$ is obtained. Note that this is a pessimistic estimate, as G has typically fewer than c^2 edges.

5.3 Simplification Based on the Comparison Measure Contour Tree (CTS)

Simplification based on the reachability graph has a major drawback. If the number of elements $\mathbb{E}(f)$ is relatively large, the simplification algorithm discussed in the previous section is costly, since the ascending and descending sets must be calculated for every component.

The edge weight $w(S, T)$, based on the opening angle $\alpha_\sigma^+(x)$ can however be used to modify a simplification approach presented by Suthambhara and Natarajan [121] for the Jacobi set of two scalar functions such that it works for the Pareto set for arbitrarily many functions.

Suthambhara and Natarajan calculate a local comparison measure that reflects local topological difference between the functions at each point $x \in \mathbb{S}$. This measure yields a scalar field f' from which a contour tree [32] is extracted. Each node $v \in V$

corresponds to connected components in the *level sets*: $\{x \in \mathbb{S} \mid f'(x) = c_v\}$ for some value c_v . Each edge (a, b) corresponds to a connected region in

$$\{x \in \mathbb{S} \mid c_a \leq f'(x) \leq c_b\}$$

such that the level sets for the nodes a and b are subsets. Suthambhara and Natarajan then proved that the hypervolume of this connected region is a conservative estimate of the change in relationship between the two functions if specific graph operations are applied to the contour tree.

Therefore their approach is extended by introducing a similar local comparison measure with the same interpretation as in the work of Suthambhara and Natarajan. Here, this produces a second method to simplify Pareto sets with reduced computational effort. In the following Section 5.4, both simplification techniques are compared, among others to provide multiple, different points of views on multifield simplifications and also to indicate some initial indications of a connection between Jacobi and Pareto sets. This connection is discussed in detail in Section 6.2.

5.3.1 Comparison Measure

Suthambhara and Natarajan's comparison measure is defined for every point in the domain. It is a scalar invariant value that is also invariant to specific triangulations of the domain, i.e. the size of the simplices. It reflects local topological differences between the underlying functions. In the context of the Pareto set approach, strong differences between the functions result in incomparable points, reducing the number of comparable ones and produces Pareto optima.

Hence, it is reasonable to replace the measure from Suthambhara and Natarajan [121] with $2 \cdot (\pi - \alpha_\sigma^+(x))$, the function from the previous section which reflects an upper bound for the difference between the separate functions in f .

As discussed before, α_σ^+ is not defined for every point in \mathbb{S} since σ is ambiguous for some x , especially vertices. Following the mentioned work [121], α^+ is averaged at the vertices based on the values in the adjacent d -simplices. Through linear interpolation using these average values at the vertices, a continuous, piecewise linear scalar field κ_f is received for \mathbb{S} .

Note that for the Pareto optimal points for which no cone exists, κ_f equals π , which directly implies that Pareto optima are maximal in κ_f . It is easily enforced that every element in $\mathbb{E}(f)$ contains at least one Pareto optimum by adding new functions.

Assume that an element in $\mathbb{E}(f)$ does not contain a Pareto optimum and therefore only contains Pareto maxima and/or Pareto minima. For such a Pareto maximum x with $H_\sigma^-(x) \neq \{x\}$, d new functions are added to f . Every new function is equivalent to f_0 except for a sufficiently small neighborhood $U(x) \cap H_\sigma^-(x)$ around x . Inside that region, the positions with the function values $f_0(x)$ in the new functions are moved by a sufficiently small ϵ such that these positions form a small d -simplex. The other values for the new functions in $U(x) \cap H_\sigma^-(x)$ are set through linear interpolation. Note that for every point inside that simplex and for every direction from that point,

the function value increases while it decreases in another new functions. Hence, every point inside that simplex is surrounded by incomparable points and is therefore Pareto extremal.

However, since x is in the connected component and the new Pareto extrema are all adjacent to x , the new functions do not increase $|\mathbb{E}(f)|$ and at most enlarge the component including x . Also, they do not change κ_f drastically and only inside the small neighborhood $U(x)$.

Hence, for this work, it can be assumed that every element of $\mathbb{E}(f)$ contains at least one maximum with respect to κ_f . Furthermore, since all Pareto optima have π as value of κ_f , each element in $\mathbb{E}(f)$ can be associated with a unique connected component in the level set $\{x \in \mathbb{S} \mid \kappa_f(x) = \pi\}$.

Note that since the new functions are almost equal to f_0 , they can be neglected for the calculation of κ_f except for small neighborhoods around some Pareto optima where those function values differ from f_0 . Thus, only little computational effort is required to create those new functions.

5.3.2 Contour Tree

Given κ_f , the approach in this thesis follows Suthambhara and Natarajan to construct and simplify a contour tree. For the purpose of illustration, Figure 5.4 presents the graph operations provided in work by Suthambhara and Natarajan [121]. The larger blue nodes are associated with connected components in the Jacobi set. Note the similarity of the Merge and Purge operations to the graph simplification applied in Section 5.2.

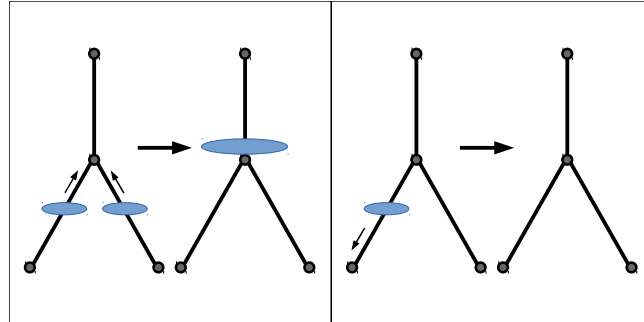


FIGURE 5.4: Illustration of the merge and purge graph operations similar to the work by Suthambhara and Natarajan [121].

Also note that, under the above assumption, all connected components are associated with leafs in the contour tree such that a merge tree is sufficient for simplification purposes and simpler to compute. Using this insight, as well as the fact that all nodes of interest, i.e. those associated elements of $\mathbb{E}(f)$, are maxima in the merge tree, a simplified version of the greedy algorithm [121] is applied. This algorithm sequentially chooses the graph operation with minimal cost, applies operations to recalculate the weights and repeats this until the minimal cost is above a given threshold or until a number of operations is reached. However, the basic idea to merge edges with smallest weight first, as is also done in the previous section, are

adopted. Here, for each edge the integral of κ_f over the area swept by the level sets corresponding to an edge in the merge tree are taken as edge weights.

5.4 Comparison and Discussion of the Simplification Approach

The two strategies, reachability graph simplification (RGS, cf. Section 5.2) and simplification based on the contour tree of κ_f (CTS, Section 5.3) are applied to two data sets to provide a proof-of-concept and an initial comparison between these methods. Computation times for the results shown in the following are in the range of seconds to minutes depending on the size of $\mathbb{E}(f)$, as discussed in Section 5.2.

5.4.1 Analytical Examples

The first data set is a synthetic example. A 4×2 set of positions $A_i = \{a_{i,j} \in [0, 1]^d | 0 \leq j \leq 4\}$ and $B_i = \{b_{i,j} \in [0, 1]^d | 0 \leq j \leq 4\}$ for $1 \leq i \leq 4$ is selected with the restriction that $0 < |a_{s,j} - a_{t,j}| < 0.1$ and $0 < |b_{s,j} - b_{t,j}| < 0.1$ for $1 \leq s, t, j \leq 4$. The functions f_i are defined as

$$f_i(x) := \sum_{j=0}^4 e^{-|a_{i,j}-x| \cdot 2} - \sum_{j=0}^4 e^{-|b_{i,j}-x| \cdot 2},$$

and combined to yield $f(x) = (f_1(x), \dots, f_4(x))$. The explicit values for f are calculated at the vertices of a regularly triangulated grid as input to the algorithms.

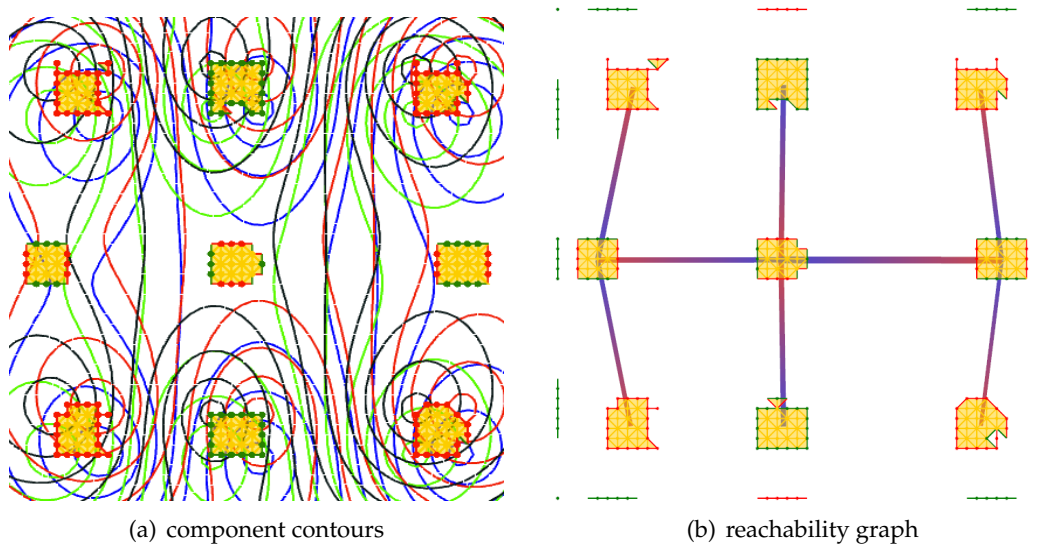


FIGURE 5.5: An illustration of the reachability graph over a synthetic data set with four scalar fields. (a) depicts the contours of each of the four scalar component functions. The reachability graph, given in (b).

Figure 5.5 provides an overview of this example. Each function has a maximum in each of the four connected components of $\mathbb{P}(f)$ in the corners, a minimum in each of the connected components in the top and bottom middle, and a saddle in each of the three connected components in the middle row. Figure 5.5(a) shows contour lines for the f_i , color coded in blue, red, green and black, to illustrate their single-field topologies. Figure 5.5(b) shows the same data and its reachability graph, together

with the Pareto sets of f consisting of minima (green), maxima (red), and optima (yellow). Nodes are placed in the center of gravity of the corresponding component. Note that loops are not shown, and edges that are already given transitively are removed to simplify visual understanding of this example.

This synthetic data set has a relatively simple structure. To obtain increased structural richness as a testbed for simplification, random noise is added. At each vertex and for each component, the function values are changed by a uniform random value from the interval $[-p \cdot \delta, p \cdot \delta]$, where δ denotes the range of the corresponding component function and $p \ll 1$.

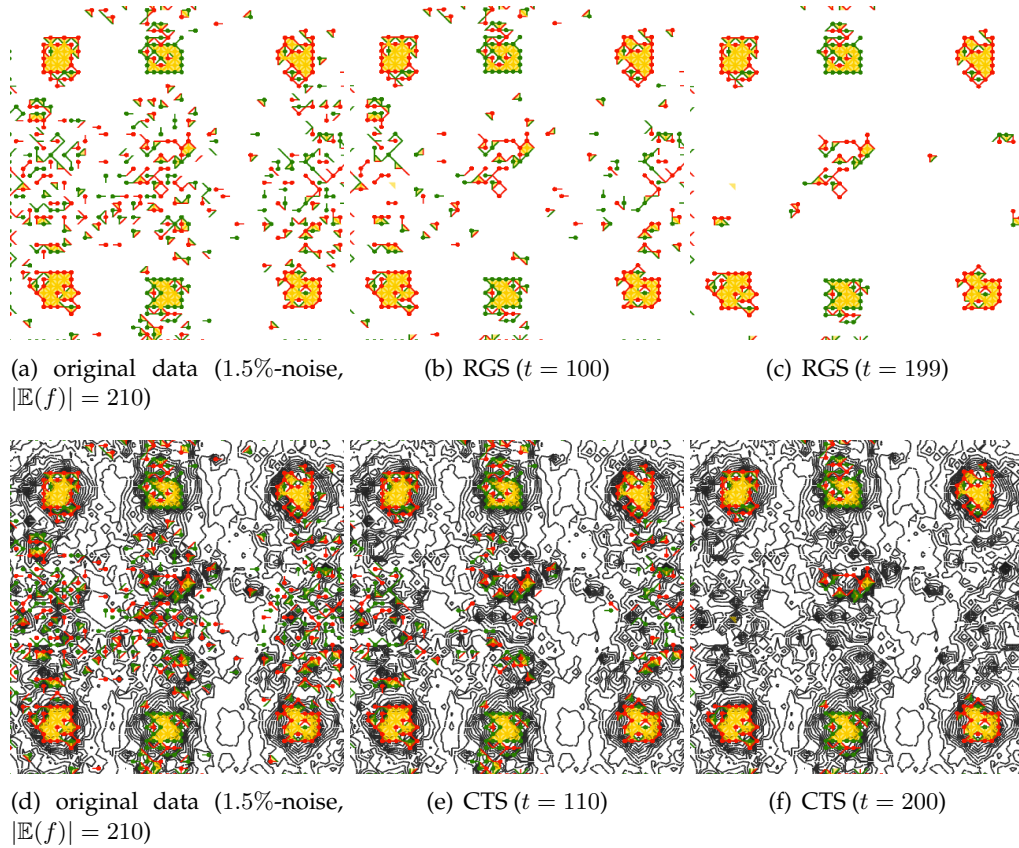


FIGURE 5.6: A comparison of the two Pareto set simplification techniques over synthetic data with 1.5% random noise. The top images (b) and (c) show reachability graph simplifications (RGS). The increasing thresholds t are indicated by the image labels. The bottom row (images (e) and (f)) illustrates the comparison measure contour tree simplification (CTS) for two similar thresholds. The first images (a) and (d) present the unsimplified data for the methods.

Figure 5.6(a) and 5.6(d) show the same synthetic data, both with additional 1.5% noise. Due to the relative flatness of f , even such a low noise level already introduces significant changes to the structure and breaks apart the three connected components in the middle into several smaller components. The six connected components in the top and bottom row remain roughly intact however. Ideally, the simplification procedure would retain these components and remove all other newly introduced components.

Figure 5.6(b) and 5.6(c), and 5.6(e) and 5.6(f) show the result of the two approaches after a number of simplification steps. In all four images, the value t indicates the number of removed, respectively merged, elements of $\mathbb{E}(f)$. In both approaches, connected components that correspond to purged or merged nodes in the reachability graph or the contour tree of κ_f , respectively, are removed. Hence, neither the form of the components nor the contour lines change among Figure 5.6(a) to 5.6(c) and 5.6(d) to 5.6(f), respectively.

Note that both methods are able to remove most of the spurious noise and keep the six connected components in the top and bottom row that are identified in the noise-free data in Figure 5.5. CTS however removed almost all of the noise-based components too, among them those which are associated with the three connected components in the middle row from Figure 5.5.

For the RGS approach, a range of simplification steps are observed for which the desired elements of $\mathbb{E}(f)$ remain. Still, an exact reconstruction of the original synthetic data, i.e. topological equality in the reachability graph, could not be achieved.

An explanation for the results is twofold. First, the induced noise results only in small, local maxima in κ_f . Hence, the Pareto extrema that are a consequence of this noise correspond to only low-weighted nodes, which are all removed very early in the CTS method. Since leafs are always removed from the contour tree the remaining nodes do not gain weight very fast. Hence, in later simplification steps, also Pareto sets which otherwise are considered important are removed. For the RGS approach, on the other hand, the ascending and descending sets of merged Pareto extrema are changed along with all edge weights adjacent to the corresponding nodes in the reachability graph. Hence, in a cluster of noise-based Pareto extrema, such as those in the middle row of Figure 5.5, it is more likely that Pareto extrema remain in the data under RGS than it is if CTS is applied.

5.4.2 Can Data

The subsection uses the Can data set (courteously provided by M. Rütten, DLR Göttingen) as a real-world example with four fields resulting from a CFD simulation of flow in a fluid-filled cylinder with a rotating lid. Section 7.1 provides a detailed overview over all individual fields as well as the general application for this kind of simulation scenario. For these examples it is sufficient to consider the Pareto set as shown in Figure 5.7 and, understand that this Pareto set provides extremal regions, i.e. elements of $\mathbb{E}(f)$, in which all fields agree. The simplicial complex that serves as input to both simplification schemes is generated as a two-dimensional slice parallel to and containing the cylinder's central axis.

Pareto maximal and minimal components, as classified by the reachability graph, are of main interest in the latter application since they indicate regions in which all considered fields suggest extremal behavior. All fields agree that towards those components their value increases or decreases, respectively, and hence agree on extremality of this region. Note that in Figure 5.7, the reachability graph is shown through lines going from red to blue in ascending direction, but without edges that are already given transitively. The motivation is to highlight the main structure of the graph, while hidden edges are still used in the simplification processes.

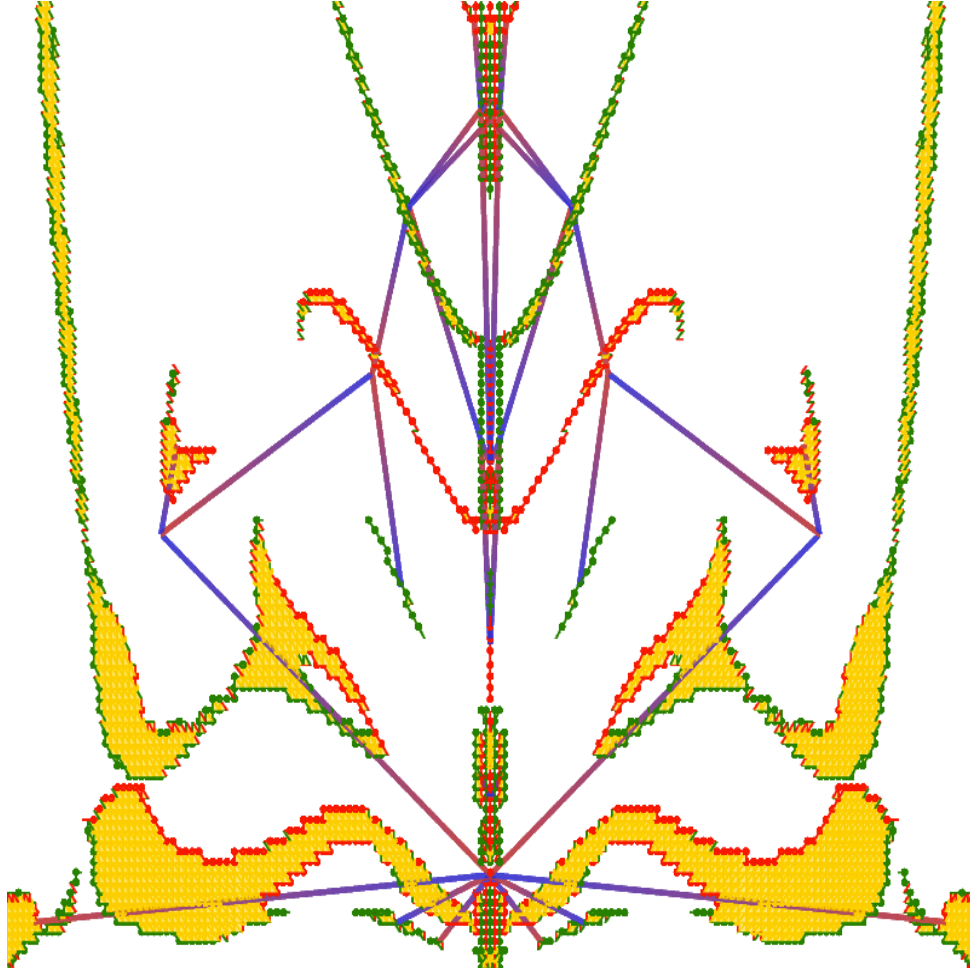


FIGURE 5.7: An illustration of the Pareto set and its reachability graph as an overlay over a real-world data set with four scalar fields, called the can data set. Note that this visualization removes transitively given edges in the reachability graph due to confusion and clutter. See Section 7.1 for more details.

Edges in the graph furthermore provide a domain expert with information on how those regions are connected or even interact with each other. The graph aids the experts to understand global structures and directs their attention to Pareto extrema, where a detailed analysis might be necessary. Specifically, an edge represents a set of paths between two connected components on which all fields agree with respect to the ascending and descending direction, respectively.

It can be observed that this data set is relatively clean and contains little structure. Again, noise is added to gauge the effect of simplification. Figure 5.8(a) and 5.8(d) present the Pareto set for modified data with random 0.04-noise. Note how such a small degree of noise already produces a large amount of separate extremal regions.

Figure 5.8(b) and 5.8(c) present the results of RGS for two different numbers of simplification steps, whereas Figure 5.8(e) and 5.8(f) show the outcome of the CTS

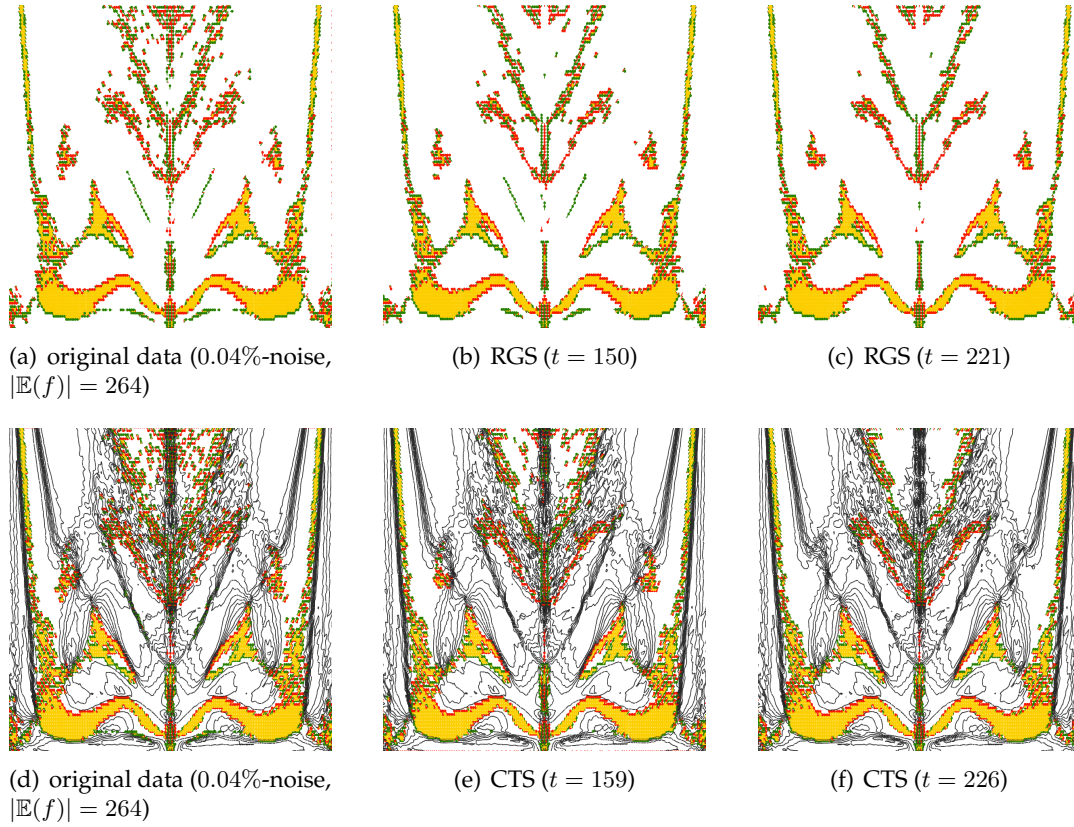


FIGURE 5.8: A comparison of the two Pareto set simplification techniques over synthetic data with 0.04% random noise. The top images (b) and (c) shows reachability-graph simplification (RGS) for two increasing thresholds. The bottom row illustrates the comparison measure contour tree simplification (CTS) for two increasing thresholds (images (e) and (f)). The unsimplified data is presented in the images (a) and (d), respectively.

method. The value t indicates the number of removed, respectively merged, connected components of $\mathbb{P}(f)$.

As with the previously investigated, synthetic data, it is noticeable that the contour tree approach removes more components than necessary, while in the reachability graph approach noise-based extremal regions remain.

5.4.3 Discussion

In this chapter, two techniques for simplifying multifield scalar data based on structures gained through the calculation of the Pareto set are presented.

The first technique is based on the reachability graph – a weighted graph whose nodes represent connected components of Pareto extrema. Connectivity is determined by the existence of paths along which function values that are strictly increasing or decreasing. Edge weights estimate the local stability as measured by the opening angle of the space of incomparable points within a region. For two fields, this measurement is a normalized version the approach presented by Suthambhara

and Natarajan [121], originally developed for Jacobi set simplification. However, these presented version extends naturally to more than two fields and is furthermore resilient to scaling input field values, e.g. resulting from a change in units. The simplification itself greedily merges or purges node and pairs of nodes, respectively, in the order of edge weights until a given threshold is reached. The concept of reachability also lends itself to a decomposition of the domain by defining an equivalence relation among regular points from which the same sets of Pareto extrema can be reached.

The second simplification technique follows more closely along the lines of the idea introduced by Suthambhara and Natarajan [121]. It computes the contour tree of the stability field computed from the original multifield data and performs topological simplification on that contour tree. Again the stability measure of Suthambhara and Natarajan is replaced by the opening angle measurement to support more than two fields.

The results of a case-study-based qualitative comparison of both methods revealed that both techniques manage to remove most of the artificial noise that is introduced during the experiments, but the contour-tree-based method removes more connected components than necessary. The different results may be explained by observing that the CTS essentially turns the multifield problem into a single-field problem, using only local information at each point, whereas the reachability graph considers the functions jointly within regions and thus has more information at its disposal. Compared to the contour tree technique, the reachability graph supports loops and can thereby handle errors based on inconvenient triangulation. However, loops remain difficult to interpret, because they never arise in the single-field setting. In addition, the contour tree technique is slightly faster, but neither does it provide a good abstract view on the data as the reachability graph nor the above indicated domain decomposition. However, these observations are based on a few scenario cases, so that potential advantages for both approaches could not yet be identified.

All presented simplification methods, including the one by Suthambhara and Natarajan, currently lack precise rules to implicitly change the scalar values of the multifield. Hence the combinatorial structures, i.e. the contour tree and the reachability graph, respectively, are simplified instead. Suthambhara and Natarajan point out that the corresponding modifications of the multifield itself may be complex and computationally expensive since it must satisfy numerous constraints on the local stability measure. However, the feasibility of such modifications for the *RG* and the Pareto set are outlined in this thesis. It presents three concrete models to change the multifield data set corresponding to a simplification of the reachability graph, but further analysis of these models, for example their efficiency, runtime and limitations, may depend on the actual application.

Although only two-dimensional domains were considered in this chapter, the computation and simplification of the combinatorial structures extend naturally to higher dimensions and can always be shown via graph drawing. To better visualize the simplification, in detail why and where simplification is applied, it is necessary to show the intersection between ascending and descending sets of the connected components of the Pareto set, corresponding to merged nodes. This is, however, challenging and possible directions for future work need methods like multifield

transfer function design for volume rendering on the combinatorial structure in analogy to Zhou and Takatsuka [192].

While in this chapter the theoretical feasibility of Pareto set simplification is demonstrated, a more in-depth evaluation is desirable. Furthermore, the presented methods are limited by the computational overhead to create the support structures, especially the reachability graph. The creation of the graph is extremely time consuming in contrast naive Gaussian smoothing in case of a reasonable large number of fields and so a major obstacle for application purposes. In part, this is due to the fact that for each component in $\mathbb{E}(f)$ all existing ascending paths have to be traced.

However, it is possible to find relations between Pareto and Jacobi sets, as well as other topology-based techniques. This similarity can be used to transform topological simplification for these concepts to the Pareto set concept as done with Jacobi set simplification [121] in the CTS section. Similar benefits may arise through a comparison with Morse-Smale complexes [76] and joint contour nets [30], especially how the quantization of the fields' values interacts with the reachability relation. Because ascending and descending paths are quite unlikely, steepest ascent and descent the relation to Morse-Smale complexes is not imminent. But Szymczak's work [156] on piecewise constant vector fields bears some algorithmic similarity to the detection of Pareto sets and may provide a link between the two.

Chapter 6

Comparison of the Pareto set with Similar Multifield Techniques

In this chapter the Pareto set concept is compared with existing, related techniques, both in terms of limitations and functionality. This places the Pareto set in the context of the general topic of topology-based multifield analysis and visualization tools. Hence, in this chapter, advantages and drawbacks of the most common topology-based approaches are discussed as well as their limitations compared to the Pareto set. This helps domain experts to better choose which technique is best suited for their scenario and to better interpret differences in results and visualizations.

Another, major segment of this chapter is the comparison the techniques in terms of transformation. In detail, in this thesis, equality or subset relation between the different concepts are presented to integrate existing algorithms and ideas from previous and future work into the Pareto set concept and visa versa. Thus, improving the calculation and implementation of the Pareto set and its related structures, as well as following simplification approaches now and in the long run.

6.1 General Comparison with Common Methods

To provide a general overview and context, the Pareto set is compared with related topology-based definitions for multiple scalar fields. Because the presented methods differ structurally, the general comparison focuses on presenting features, limitations, and properties that distinguish the different approaches from each other, so that prospective applicants can decide which method best fulfills their needs. This sections starts with presenting the related methods and some of their unique characteristics and finishes with common features and differences of the Pareto set approach to these methods. A more detailed comparison, especially with the focus on how results from a particular method can be used to calculate structures from the Pareto set concept and vice versa, can be found in Section 6.2 and 6.3. The general comparison presented in the current section can also be found in a manuscript that was published in *Computer Graphics Forum* [91].

Jacobi Sets: Edelsbrunner and Harer [52] defined the Jacobi set of Morse functions as the set of critical points of one function's restrictions to the intersection of the other functions' level sets augmented by the critical points of the single functions. Although not immediate from the definition, Jacobi sets do not depend on the order in which the fields are given. Equivalently, a point x is in the Jacobi set, if the Jacobi matrix, i.e. the matrix of the m gradients at x , has a rank lower than m . This highlights that the method is only applicable as long as the number of fields does

not exceed the dimension of the domain manifold. The Jacobi set concept is parameter free, i.e. their calculation does not need further input except the multifield data, has been used to define time-varying Reeb graphs of single scalar fields on a 3-sphere [57], and culminates in the definition of Reeb spaces [53].

Comparison of Largest Contours: Schneider et al. [145] in an extension of earlier work [146] proposed the following approach: for each given scalar field compute its largest contours, i.e. maximal contours that contain only one critical point, then compute similarities between largest contours of all fields based on a normalized spatial overlap measure and store this information in a weighted graph to which graph clustering is applied. Largest contours are considered to be volumetric features. Therefore, graph clusters can be interpreted as sets of features that are consistent across fields. The user can tune the result by selecting and deselecting clusters and single largest contours in a gallery view. The similarity graph and its clustering are no typical topological structures and the information is not integrated into the domain, e.g. to give a subdivision. The computation of largest contours requires the computation and topological simplification of a contour tree for each field, limiting the method to simply-connected domains and requires parameter input to guide the simplification.

Joint Contour Nets (JCN): A recent approach is presented by Carr and Duke [30]. It splits each cell into so-called slabs, which are connected regions of points equal in quantized/rounded multifield values, then generates a graph that encodes the adjacency of these slabs. This graph represents the multifield topology for the chosen quantization and is presented using graph drawing techniques. Visual inspection can reveal structures resembling local extrema and saddles. The method requires quantization of field values, requiring a quantization parameter per field. However, these parameters can be used to drive topological simplification. From the joint contour net the quantized contour trees of the single fields can be extracted using a simple quotient graph algorithm.

Pareto set: Both methods, Pareto sets and JCNs, work for any dimensions and on any number of scalar fields. Jacobi sets, however, have not been discussed for a case in which the number of fields exceeds the number of dimensions of the manifold domain. Although conceptually work for any discretization of the domain, all algorithms based on the concepts discussed in this section restricted their implementation to simplicial complexes and piecewise linear data.

Unlike the other methods, Pareto sets are not blind to inverting one or more of the input fields, which means that changing field's gradient signs affects the outcome. This implies that the Pareto set concept requires the applicant to invert fields meaningfully, e.g. when it is reasonable to assume that local maxima of one field semantically match local minima of another, the applicant has to invert one of the two. As a small example, consider the identification of "bad weather" with a temperature and a precipitation map. Usually, bad weather consists of low temperatures and high precipitation. Hence, if, e.g. the temperature map inverted, bad weather is located where both variables are maximal.

All methods can be considered extensions of "classical" topological structures: Jacobi sets extend the notion of critical points, Schneider et al. [145]'s approach extends largest contours, and Joint contour nets extend contour trees. While the Pareto

set extends the notion of local extrema of single scalar fields to Pareto optima in multifields, it does not extend the notion of saddles, although saddle-like features tend to show up in the results. The Pareto set concept extends the notion of monotone paths to ascending and descending paths. The connectivity of Pareto extremal regions is determined by observing ascending and descending sets between them and encoding them into graphs as done in the previous section. Non-Pareto extremal points are defined as equivalent if their ascending and descending sets border the same Pareto extremal region. A decomposition of the domain using this equivalence relation and then considering neighborhood relations between regions can also reveal interesting aspects of the data. For the special case of one scalar field on a simply connected domain, the reachability graph is precisely the contour tree, as is shown by Chiang et al. [38], who use monotone path to define and compute the contour tree.

6.2 Comparison of the Pareto set with the Jacobi set

In this section deeper connections between Pareto sets and Jacobi sets [52] are examined, specifically, to prove a subset relation between the sets of critical points defined by both concepts. The results indicate further relations between both sets and other multifield approaches, and, in particular, the Morse decomposition and the JCN methods. All results of this section can also be found in previous work [87].

6.2.1 Alternative Definition of Pareto Sets based on Linear Programs

To show the relation between the Jacobi set and the Pareto set, first, an alternative characterization of critical points in a piecewise linear multifield through point-specific linear programs is defined. Then it is proven that a point is in the Pareto set if and only if the corresponding linear program has no solution, i.e. is infeasible, thus showing that both definitions are equivalent. This allows mathematical theorems from linear programming to be applied for the proofs in Subsection 6.2.5.

To identify the Pareto set, an intersection of halfspaces characterized by the gradients of f is calculated. Hence it is reasonable that the optimization problem should be based on a suitable description of the latter. Note that for each point $x \in \mathbb{M}$ those gradients are only related to a sufficiently small neighborhood $U(x)$. Hence, for each point x , each Morse function f_i can be locally approximated by a linear function g_i with the only condition that $g_i(x) = f_i(x)$ and $\nabla g_i(x) = \nabla f_i(x)$. With this function g_i , $H_{i,U(x)}^+(x) = \{x\}$ is equivalent to $H_{i,\mathbb{M}}^+(x) = \{x\}$ when f_i is substituted by g_i . Note that g_i is different for each x and not a global approximation of f_i . For easier reading, the $U(x)$ in $H_{U(x)}^+(x)$ is dropped in the following theorem.

Theorem 6.2.1. *Let $\varepsilon > 0$. Given a point $p \in \mathbb{M}$, and denote $k_i := \nabla f_i(p)$.*

The $(n+1) \times (d+1)$ -matrix A_p is defined as

$$A_p := \begin{bmatrix} -k_{1,1} & \cdots & -k_{1,d} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ -k_{n,1} & \cdots & -k_{n,d} & 1 \\ 0 & \cdots & 0 & -1 \end{bmatrix}$$

and vectors

$$\begin{aligned} b &= (-\varepsilon, \dots, -\varepsilon, 0)^T \in \mathbb{R}^{n+1} \\ c &= (0, \dots, 0, 1)^T \in \mathbb{R}^{d+1}. \end{aligned}$$

Then $H^+(p) \neq \{x\}$ if and only if the linear program \mathcal{P}_p^+ defined by

$$\begin{aligned} &\text{maximize} && c^T x \\ &\text{subject to} && A_p x \leq b \end{aligned} \tag{6.1}$$

is feasible and unbounded. Feasible means that an $x \in \mathbb{R}^{d+1}$ exists for which $A_p x \leq b$ holds and unbounded means that for every x with $A_p x \leq b$ an $x' \neq x$ with $A_p x' \leq b$ can be found such that $c^T x < c^T x'$.

Proof. Without loss of generality, it is assumed that $p = 0$ by translation such that through the introduction of $\varepsilon > 0$, p is not a solution. Otherwise, without ε , i.e. with $b = \vec{0}$ the zero vector, the linear problem would always be feasible with $x = \vec{0}$ as a solution.

The two directions of the equivalence claim are proven separately. For \Rightarrow , $H^+(p) \neq \{x\}$ is assumed such that $y \in H^+(\vec{0})$ with $y \neq 0$ exists. Then, based on the definition of $H^+(\vec{0})$, $(y - \vec{0})^T \cdot k_i > 0$ for all $i = 1, \dots, n$. Hence, $\varepsilon_i > 0$ exists such that $k_i^T \cdot y \geq \varepsilon_i$ for all $i = 1, \dots, n$. Thus, for $\varepsilon_{\min} = \min_{i=1, \dots, n} \varepsilon_i$ and a $\gamma_0 \geq \frac{\varepsilon}{\varepsilon_{\min}} > 0$,

$$k_i^T \cdot y \cdot \gamma \geq \varepsilon_{\min} \cdot \gamma \geq \varepsilon \text{ holds for all } i = 1, \dots, n \text{ and any } \gamma \geq \gamma_0.$$

Note that an arbitrary γ is introduced in this proof to receive an infinite amount of solutions for \mathcal{P}_p^+ . To find these solutions, for any $\gamma \geq \gamma_0$, a $q \geq 0$ is set such that the following inequality holds for all $i = 1, \dots, n$:

$$(k_i^T \cdot y \cdot \gamma) - q \geq \varepsilon$$

This is true as long as q lies within the difference of $k_i^T \cdot y \cdot \gamma$ and ε , i.e.

$$0 \leq q \leq (y \cdot k_i^T \cdot \gamma) - \varepsilon, \text{ for all } i = 1, \dots, n.$$

Hence, for $q := \min_{i=1, \dots, n} (k_i^T \cdot y \cdot \gamma) - \varepsilon$, it is easily computed that for $y' = (y_0, \dots, y_n, q)$

$$A_p \cdot y' \leq b \text{ holds.}$$

To see this, consider the calculation of line i , $1 \leq i \leq n$, with $A_{p,i} \cdot y' \leq -\varepsilon$. Given the definition of A_p this calculation is equivalent to $-k_i^T \cdot y + q \leq -\varepsilon$ and thus, is simply an inversion to the above argumentation, everything is multiplied by -1 . Due to the inversion the parameter q changes into $-q \leq 0$ such that the last line, $A_{p,n+1} \cdot y' \leq 0$ is also ensured. Hence \mathcal{P}^+ is feasible. Furthermore, with $c^T y' = q = \gamma \cdot \min_{i=1, \dots, n} (k_i^T \cdot y) - \varepsilon$ and $\gamma \geq \gamma_0 > 0$ not bounded from above, the program is also unbounded.

Conversely for \Leftarrow , let $y \in \mathbb{R}^{n+1}$ be a solution to the linear program \mathcal{P}_p^+ . Define $y' = (y_0, \dots, y_n)$ such that, based on the definition of \mathcal{P}_p^+ ,

$$-k_i^T \cdot y' + y_{n+1} \leq -\varepsilon \text{ and therefore } k_i^T \cdot y' \geq \varepsilon + y_{n+1} \text{ hold.}$$

Hence, with $A_{p,n+1} \cdot y = -1 \cdot y_{n+1} \leq 0$, it holds that

$$k_i^T \cdot y' > 0$$

for all $i = 1, \dots, n$, such that $y' \neq \vec{0}$ is in $H^+(\vec{0})$ by definition of the k_i . \square

The linear program \mathcal{P}_p^- for which the gradient vectors are negated is defined analogously, and its relation to $H^-(p)$ is proven in the same way. Also note that, as a direct implication of Theorem 6.2.1, $p \in \mathbb{M}$ is not Pareto maximal, i.e. $p \notin \mathbb{P}^+$ if and only if \mathcal{P}_p^+ is feasible and unbounded.

6.2.2 Definition of Jacobi Sets based on Linear Programs

In the following subsection, the concept underlying the Jacobi sets is briefly revisited. A more thorough elaboration can be found in related work by Edelsbrunner [52].

6.2.3 Mathematical Background of Jacobi set

The Jacobi set of $n \leq d$ Morse functions f_i on a common d -manifold \mathbb{M} is defined through level set intersections. The *level set* of a function f_i , $1 \leq i \leq n$, and scalar value $t \in \mathbb{R}$ is the set $\mathbb{M}_t := f_i^{-1}(t)$. For multiple functions, the intersection of the level sets, excluding function f_l , is defined as $\mathbb{M}_{t,l} := \bigcap_{i \neq l} f_i^{-1}(t_i)$ for a vector $t \in \mathbb{R}^n$. The Jacobi set \mathbb{J} is the closed set of all points x that are critical in f_l restricted on $\mathbb{M}_{f(x),l}$ for some index $1 \leq l \leq n$ with respect to scalar field topology.

Note that a level set on a d -manifold is of dimension $d - 1$ and the closed intersection of $n - 1$ level sets is of dimension $d - (n - 1)$.

Conversely, $x \in \mathbb{J}$ holds if and only if, after removing all points from \mathbb{M} with different values than x with respect to $n - 1$ of the Morse function, x is critical with respect to the $(n)^{th}$ function, thus a minimum, maximum or saddle of some degree. Note that x can become isolated in $\mathbb{M}_{f(x),l}$ is also considered as critical in this thesis. In case of $n > d$, this results in $\mathbb{J} = \mathbb{M}$.

In another approach towards Jacobi sets presented by Edelsbrunner [52], criticality of a point $x \in \mathbb{M}$ is defined through the gradients of the functions f_i at x . In detail, the gradients ∇f_i of the function $i \neq l$ at a point $x \in \mathbb{M}_t$ are considered for some index $1 \leq l \leq n$ and a n -vector t . Those gradients span the linear subspace of vector normals of \mathbb{M}_t . If ∇f_l at point x also belongs to this linear subspace, x is critical in f_l restricted on $\mathbb{M}_{t,l}$.

In general, Jacobi sets can be described as the closure of a set of points x for which there is a $\lambda \in \mathbb{R}^n$ such that $\lambda \neq 0$ and $\sum_{i=1}^n \lambda_i \cdot \nabla f_i(x) = \vec{0}$. Hence,

$$\mathbb{J} := \text{cl} \left\{ x \in \mathbb{M} \mid \exists \lambda \in \mathbb{R}^n \setminus \{0\} \text{ s.t. } \sum_{i=1}^n \lambda_i \nabla \cdot f_i(x) = \vec{0} \right\}$$

Note that in this linear combination negative parameters λ_i are allowed, thus the direction of the gradient is ignored.

For further use, the notation of the *positive Jacobi Set* \mathbb{J}^+ is introduced. It includes the additional restriction that $\lambda_i \geq 0$ for all $1 \leq i \leq n$ and thereby results in a positive linear combination.

Definition 6.2.1. A point $x \in \mathbb{M}$ is in the positive Jacobi Set \mathbb{J}^+ iff

$$\exists \lambda \in \mathbb{R}^n \setminus \{\vec{0}\} \text{ s.t. } \sum_{i=1}^n \lambda_i \cdot \nabla f_i(x) = \vec{0} \text{ and } \forall_{i=1}^n \lambda_i \geq 0.$$

In this work $\mathbb{J}^+ \subseteq \mathbb{J}$ is obtained as a subset of the Jacobi set. In an analogous fashion, also the notation of negative Jacobi sets \mathbb{J}^- is added. Note that both sets in the definition of \mathbb{J}^+ and \mathbb{J}^- do not include the closure. This avoids complications in the following proofs with points that lie in the closure but not in the actual set.

Furthermore, note how the alternative definition is based on the gradient vectors rather than on the values of the Morse functions. Thus, the translation of this definition for multiple scalar fields to multiple vector fields is immediate if each vector field is equivalent to a gradient fields of some Morse function.

For the following part, it is noteworthy that Equation 6.2 is closely related to the definition of critical triangles by Szymczak [156], defined for k piecewise constant vector fields over a 2-dimensional manifold. Szymczak calls a triangle critical iff the convex hull of its vectors – inside a triangle the separate vector fields are constant – contains the zero vector. Definition 6.2 and the relation to Jacobi sets are presented to use the results in the main proof and, additionally, to put Pareto sets transitively in relation with Szymczak's definition.

$$\mathbb{C}^+ := \left\{ x \in \mathbb{M} \mid \exists \lambda \in \mathbb{R}^n \text{ s.t. } \sum_{i=1}^n \lambda_i \cdot v_i(x) = \vec{0} \wedge \sum_{i=1}^n \lambda_i = 1 \wedge \lambda_i \geq 0 \text{ f.a. } i \right\} \quad (6.2)$$

Claim 6.2.2. Under the assumption that for every $x \in \mathbb{M}$, $v_i(x) = \nabla f_i(x)$ with f_i the Morse functions as given above, it holds that $\mathbb{C}^+ = \mathbb{J}^+$.

Proof. \subseteq : Obviously, if a convex combination with parameter λ results in the zero-vector, those parameters are also a positive linear combination that result in the zero-vector. Furthermore, $\sum \lambda_i = 1 \wedge \lambda_i \geq 0$ f.a. i excludes $\lambda = 0$ as parameter.

\supseteq : If the positive linear combination with parameter λ results in the zero-vector, it can be concluded that $\sum \lambda_i = c \neq 0$ holds. The right inequality is implied by the requirement that $\lambda \neq \vec{0}$ and $\lambda_i \geq 0$ for $0 \leq i \leq k$. Hence, with the new parameter $\lambda' = \lambda/c$ the sum $\sum \lambda_i = 1$ is received and the convex combination still results in the zero-vector. \square

6.2.4 Alternative Definition of Jacobi Sets based on Linear Programs

As with the Pareto sets in the previous subsections, in this subsection, points in the Jacobi set are characterized through the feasibility of a linear program. This translation allows the usage of mathematical results from a linear program in the proof in Subsection 6.2.5 and furthermore provides a new approach to compute Jacobi sets.

Theorem 6.2.3. Let again $\varepsilon > 0$, $p \in \mathbb{M}$, and $k_i = \nabla f_i(p)$ to define a $(d+1) \times (n+1)$ -matrix

$$B_p := \begin{bmatrix} -k_{1,1} & \cdots & -k_{d,1} & 0 \\ \vdots & \ddots & \vdots & \vdots \\ -k_{1,d} & \cdots & -k_{m,d} & 0 \\ 1 & \cdots & 1 & -1 \end{bmatrix}$$

and vectors

$$\begin{aligned} b &= (-\varepsilon, \dots, -\varepsilon, 0)^T \in \mathbb{R}^{n+1} \\ c &= (0, \dots, 0, 1)^T \in \mathbb{R}^{d+1} \end{aligned}$$

Then $p \in \mathbb{J}^+$ if and only if the linear program \mathcal{D}_p^+

$$\begin{aligned} &\text{minimize} && b^T y \\ &\text{subject to} && B_p y = c \\ &&& \text{and} \quad y_i \geq 0 \end{aligned} \tag{6.3}$$

is feasible and unbounded.

Proof. For \Rightarrow , let without loss of generality $p = \vec{0}$ by translation and assume that $p \in \mathbb{J}^+$. By equation 6.2.1 and the proven equality $\mathbb{J}^+ = \mathbb{C}^+$, there is a $\lambda \in \mathbb{R}^d$ such that

$$\sum_{i=1}^n \lambda_i \cdot k_i = \vec{0} \quad \text{with} \quad \lambda_i \geq 0 \quad \text{and} \quad \sum_{i=1}^n \lambda_i = 1$$

holds. Thus, a possible solution for \mathcal{D}_p^+ with arbitrary $\alpha > 0$ is given by

$$y := (\alpha \lambda_1, \dots, \alpha \lambda_d, (\alpha - 1))^T$$

Since $\sum_{i=1}^n \lambda_i k_i = \vec{0}$ holds, also $\sum_{i=1}^n \lambda_i \alpha k_i = 0$ is true such that the first d rows of $B_p y = c$ are true. With

$$\left(\sum_{i=1}^n y_i \right) - y_{(n+1)} = \left(\sum_{i=1}^n \lambda_i \right) \alpha - (\alpha - 1) = 1$$

all conditions of \mathcal{D}_p^+ are fulfilled by y which is therefore feasible. Furthermore, since

$$b^T y = -\varepsilon \cdot \alpha \cdot \sum_{i=1}^n \lambda_i = -\varepsilon \cdot \alpha$$

can be arbitrary small with respect to α , \mathcal{D}_p^+ is also unbounded.

Conversely for \Leftarrow , assume that \mathcal{D}_p^+ is feasible with $y \in \mathbb{R}^{n+1}$ as a possible solution. First, define $\alpha := y_{(n+1)} + 1$ and parameter vector $\lambda \in \mathbb{R}^n$ with

$$\lambda_i = y_i / \alpha \quad \text{for all } 1 \leq i \leq n.$$

The problem condition $y_i \geq 0$ holds for all i especially $i = n+1$ that implies $\lambda_i \geq 0$ for

all i . This and the last row of the equation system $B_p \cdot y = c$, namely $\sum_{i=0}^n y_i - y_{n+1} = 1$, furthermore imply that $\sum_{i=0}^n \lambda_i = 1$ holds. Also, based on the first d rows of B_p ,

$$\left(\sum_{i=1}^n k_i \cdot y_i = \vec{0} \right) \Rightarrow \left(\sum_{i=1}^n k_i \cdot \lambda_i = \vec{0} \right) \text{ holds.}$$

Hence, with λ as parameters, $p \in \mathbb{J}^+$ according to equation 6.2.1. \square

As usual in this thesis, a analogous linear program \mathcal{D}_p^- is defined for which an analogous statement is easily proven, namely that $p \in \mathbb{J}^-$ if and only if \mathcal{D}_p^- is feasible and unbounded.

6.2.5 Translation between Jacobi and Pareto Sets

Having characterized both Pareto extrema and Jacobi critical points in terms of the feasibility of linear programs, the main result is stated in the next subsection. The proof is based on the *weak duality* theorem for a *primal-dual* system [20]. The following segment briefly reiterates this statement and its preconditions. Therefore, given a linear program in *primal* form

$$\begin{aligned} & \text{maximize } c^T x \\ & \text{subject to } Ax \leq b \end{aligned}$$

and a corresponding problem in *dual* form

$$\begin{aligned} & \text{minimize } b^T y \\ & \text{subject to } A^T y = c \\ & \text{and } y \geq 0 \end{aligned}$$

for which the matrix A and the vectors b and c are the same, linear programming theory states the following result.

Theorem 6.2.4 (Weak Duality). *Proven by Boyd and Vandenberghe [20] in the context of linear optimization,*

- the primal program is unbounded \Leftrightarrow the dual program is infeasible, and
- the dual program is unbounded \Leftrightarrow the primal program is infeasible.

Thus, the previous results from Sections 6.2.1 and 6.2.2 are used to easily prove the following result. This is possible since \mathcal{P}_p^+ , \mathcal{P}_p^- , \mathcal{D}_p^+ , and \mathcal{D}_p^- are all in primal and dual form, respectively.

Theorem 6.2.5. *Given a d -manifold $\mathbb{M} \subseteq \mathbb{R}^d$ and n Morse functions $f_i : \mathbb{M} \mapsto \mathbb{R}$ with $n \leq d$, the following relations are obtained:*

- (i) $\mathbb{P}^+ = \mathbb{J}^+$
- (ii) $\mathbb{P}^- = \mathbb{J}^-$
- (iii) $\mathbb{P} \subseteq \mathbb{J}$

Proof. (i). Given a point $p \notin \mathbb{P}^+$, it holds by definition of \mathbb{P}^+ that $H^+(p) \neq \{p\}$. Therefore, using Theorem 6.2.1, the linear program \mathcal{P}_p^+ is feasible and unbounded. Since \mathcal{P}_p^+ is in primal form, weak duality provides that the corresponding dual problem \mathcal{D}_p^+ is infeasible, hence neither feasible nor unbound. Thus, Theorem 6.2.3 implies that $p \notin \mathbb{J}^+$ and therefore $\mathbb{J}^+ \subseteq \mathbb{P}^+$.

The reverse statement, $\mathbb{P}^+ \subseteq \mathbb{J}^+$, follows analogously using the second line of the weak duality theorem. Combined, $\mathbb{P}^+ = \mathbb{J}^+$ is received.

Claim (ii) follows analogously by considering \mathcal{P}_p^- and \mathcal{D}_p^- instead of \mathcal{P}_p^+ and \mathcal{D}_p^+ .

(iii). The claims (i) and (ii) are used for the second equality of

$$\mathbb{P} = \mathbb{P}^+ \cup \mathbb{P}^- = \mathbb{J}^+ \cup \mathbb{J}^- \subseteq \mathbb{J}.$$

□

Note that (iii) can be trivially extended to the case $n > d$ since then $\mathbb{J} = \mathbb{M}$ and therefore $\mathbb{P} \subseteq \mathbb{J}$ as a direct result. $\mathbb{J} = \mathbb{M}$ follows directly from the level set-based definition of Jacobi sets. For each point p , after the restriction of a function to the level set intersection of the other $n - 1 \geq d$ functions, only p remains which is obviously critical under such a restriction.

Examples in the following section indicate furthermore that the subset relation (iii) is strong, i.e. $\mathbb{P} \subset \mathbb{J}$ is possible.

6.2.6 Discussion

The result obtained here is based on a continuous formulation and requires Morse functions over a d -manifold, while practical application need to work on a piecewise linear setting over d -simplicial complexes. Similar to the implementation to determine the Pareto sets in Section 4.5, for any point x on d' -simplices with $d' < d$, the linear programs for all adjacent d -simplices need to be calculated with additional conditions. Obviously, the neighborhood of the considered point x has to be restricted to the current d -simplex. This can be achieved by adding the $d - 1$ -simplices adjacent to x as new conditions to the linear program. While not proven, numerical experiments indicate that there is a similar correspondence.

Discussion

Pareto and Jacobi sets are calculated for a selection of artificial and practical data sets, given on two-dimensional simplicial meshes using the algorithms described in Chapter 4 and in the work by Edelsbrunner and Harer [52]. Figures 6.1 and 6.2 provide the results from two data sets as examples.

First, for a simulation of a fluid-filled can with a rotating bottom different criteria, here the velocity and pressure values, are measured to identify locations of possible vortices. The complete data set is presented in Section 7.1 in the next chapter. On a 2D cross-section through the can, both the Jacobi set and the Pareto set are calculated and presented in Figure 6.1. Pareto extrema are shown in red (minima) and green (maxima), respectively, while points in the Jacobi set are colored gray. Note that $\mathbb{P}_1 \subset \mathbb{J}_1$ holds.

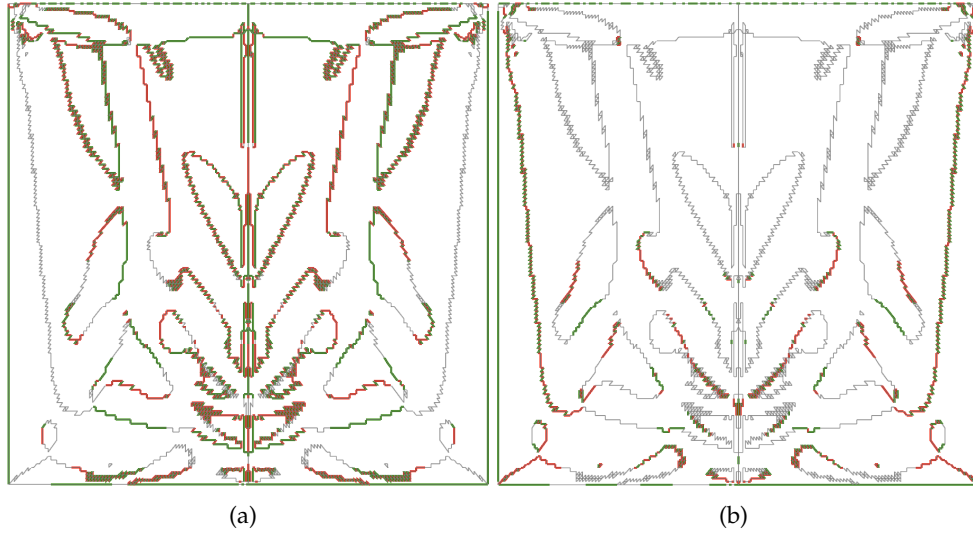


FIGURE 6.1: Pareto sets and Jacobi sets for a two-dimensional example with two functions. In both images the same fields are used but in the second image, one of the individual fields is inverted.

In Subfigure 6.1(b), one of the individual fields is negated, and the results are shown using the same coloring as in Subfigure 6.1(a); here, $\mathbb{P}_2 \subset \mathbb{J}_2$. Changing the sign of the component functions, which inverts the gradient vectors, does not change the Jacobi set since the definition of \mathbb{J} allows positive and negative parameter λ_i . Therefore, the equality $\mathbb{J}_1 = \mathbb{J}_2$ holds. In the definition of \mathbb{J}^+ and \mathbb{J}^- , the orientation of the gradients is restricted through the limitation to only positive or only negative parameters λ_i . Since other possible orientations can be neglected based on the symmetry of Jacobi sets, it follows that for the closure $cl\{\mathbb{J}_1^+ \cup \mathbb{J}_2^+\} = \mathbb{J}_1$ holds in these images. Hence, using the previous theorem, $cl\{\mathbb{P}_1^+ \cup \mathbb{P}_2^+\} = \mathbb{J}_1$ is also true. This can obviously be extended to cases with even more individual fields as long as $n \leq d$, which allows for the calculation of the Jacobi set through the calculation of the Pareto sets.

Considering the two-dimensional can data set in Figure 6.1, the points $p \in \mathbb{P}$ are those locations where the gradients of both functions point in parallel and in opposite directions. The points $p \in \mathbb{J} \setminus \mathbb{P}$ are the locations where the gradients point in the same direction. Hence $\mathbb{J} \setminus \mathbb{P}$ includes all points with parallel gradient vectors, a common direction over all fields of steepest ascent. Note how the Jacobi set connects the components in the Pareto set through one-dimensional paths. Hence, in this thesis, $\mathbb{J} \setminus \mathbb{P}$ can be interpreted as connections of the components in \mathbb{P} which form the *fastest* paths through the domain following the gradients of the individual fields. Future work towards multifield visualization and topological structures should include a study of this observation to aim for an incorporation of Jacobi and Pareto sets to build global structures and connections.

As an additional remark, note that both the Jacobi set and the Pareto set behave in a zigzag, serpentine-like manner instead of continuous lines. This nicely shows how the triangulation of the domain creates spurious elements in both sets, similar to the observations by Mascarenhas and Snoeyink [113].

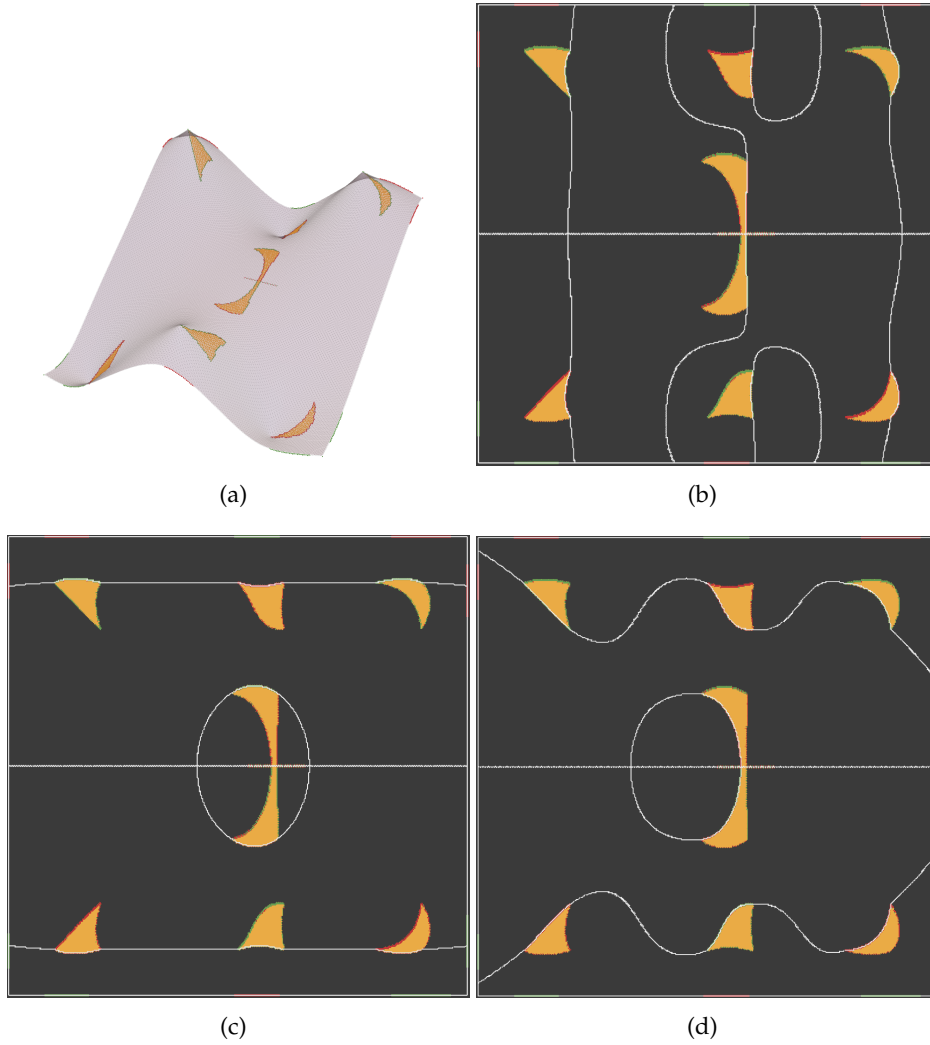


FIGURE 6.2: Pareto sets and Jacobi sets for a two-dimensional example with three similar Gaussian functions. Subfigure 6.2(a) illustrates one of the three functions as a height map. The remaining images each show the Jacobi set (white lines) for one of the three possible pairings of the Gaussian functions. All images also show the Pareto set for the multifield containing all three functions together.

The second data set consists of three artificial fields obtained from sums of Gaussians (see Section 3.3). Each individual field has three minima and maxima. The location of the minima and maxima are only slightly perturbed among the three fields. Figure 6.2(a) shows an illustration of one of the individual fields. The Jacobi sets are calculated and colored as in the previous figure. However, since Jacobi sets are only defined for $n \leq d$, the set is calculated separately for each two-pairings of the individual fields, shown in the three Figures 6.2(b), 6.2(c), and 6.2(d). The Pareto set based on all three fields is also shown in these figures with the same coloring as in Figure 6.1, but by additionally using yellow for Pareto optima.

Note that the Pareto sets for each pairing of two fields can be calculated through the intersection of the corresponding Jacobi sets and the Pareto set of the complete multifield. Hence, it is sufficient to only calculate the Pareto set once. Also note how

the Jacobi sets from all three figures, if combined, build the border of the components in the Pareto set. An assumption and additional direction of future research is that this introduces, together with the main proof, a new approach to calculate Pareto sets for a large number of fields through the unified Jacobi sets for small number of fields.

Relevance

The previous proofs prove the following three subset and equivalence relations:

- (i) $\mathbb{P}^+ = \mathbb{J}^+$
- (ii) $\mathbb{P}^- = \mathbb{J}^-$
- (iii) $\mathbb{P} \subseteq \mathbb{J}$

The previous subsection also provided the equivalence relation $cl\{\mathbb{P}_1^+ \cup \mathbb{P}_2^+\} = \mathbb{J}_1$ for the case of two individual fields. The discussed concept can, however, be extended to an arbitrary number of fields. This and the translation of the Jacobi and Pareto set definition to linear programs provide new approaches to calculate those sets.

New calculation tools and a better understanding of both sets and their relation towards each other are immediate results.

It is also hypothesized that an equality between Jacobi sets and the critical points defined in the Morse decomposition approach [156] holds, subsequently resulting in $\mathbb{J}^+ = \mathbb{C}^+$. However, Pareto sets and Jacobi sets are only two possible approaches to define criticality in multifield scalar data. There are also other methods to achieve this, as well as approaches specifically tailored to vector or tensor multifields. Thus, this chapter is an important contribution towards understanding the general concepts of multifield topology and its practical application, for example for simplification techniques.

6.3 Comparison of the Pareto Set with the Joint Contour Net

As stated in Section 4.1, the design of a fast algorithm to calculate the reachability graph is difficult, since a major obstacle is that the graph is based on continuous ascending paths between the critical regions. However, in contrast to other related path-based concepts like the Morse-Smale complex, for each point the steepest gradient does not exist in a multifield scenario but rather a range of ascending directions. Hence, instead of point to point connections, it is only possible to compute a connection between a point and a set of points, making the reachability graph calculation exponentially more complicated.

In addition, the connections have to be traced for every connected component in the Pareto set, making it prone to small local structures. Note that in multifield scenarios, local structures in each separate field are cumulative and either create large, significant structures, or simply increase the number of small, insignificant structures drastically in the combined data.

To improve the calculation of the Pareto set and its reachability graph, the relation between this concept and an approach by Carr et al. [30, 29], the Joint Contour Net (JCN), is analyzed. As stated in the beginning of this chapter, these studies are also important to place the concepts in context with each other. In contrast to the previous section, the results indicate that JCNS and the Pareto set have more in common than just a subset relation, as was the case for the Jacobi and the Pareto set. This highlights the significance of the two methods, JCN and Pareto set, by showing how different approaches and view points yield similar results and visualizations. Furthermore, in combination with the close relation between JCNs and the Reeb space, in this section a similar relation is implied between the Pareto set and a common multifield topology structure, the Reeb space.

The results are presented in several stages around the *Topology-based method in Visualization* (TopoInVis) conference series [88, 17] and the master thesis by Jan Bormann [16] inspired by that work.

6.3.1 Related Work for Joint Contour Nets

As part of Section 6.1, the Pareto set concept is compared superficially with other multifield approaches like Jacobi sets [52] and Joint Contour Nets [30, 29] in terms of domain restrictions as well as their individual advantages and drawbacks.

Chattopadhyay et al. [36] analyze the projection of the Jacobi set into the Reeb space, introducing the Jacobi structure of a Reeb space. This structure separates the Reeb space into simple components which can be measured in a scaling-invariant manner. This can directly be used as a multifield topology-based simplification approach [37].

In another direction, Tierny and Carr [161] recently use the Jacobi set to efficiently calculate the Reeb space [53]. Both Chattopadhyay et al. and Tierny and Carr present good examples of how the combination of two related concepts results in improvements, both algorithmically and in terms of understanding multifield topology.

Hence the relation between Reeb space and Jacobi set, and, in the previous section, the relation between Jacobi set and Pareto set, are already studied. To close the loop, i.e. the relation between Pareto sets and the Reeb space, the Joint Contour Net is considered as an intermediate step. The JCN is proposed as an approximation of the Reeb space [30, 118], and, at least for 2D data, it is easier to calculate than the continuous Reeb space itself.

6.3.2 Definition of the Joint Contour Net

For a brief summary on JCN, the multifield data is given as a set of n fields $f = (f_1, \dots, f_n)$ over a common domain such that $f_i : \mathcal{S} \mapsto \mathbb{R}$ are continuous, piecewise linear Morse functions with \mathcal{S} a triangulated manifold of dimension d , i.e. a simplicial complex.

Furthermore, an interval-based rounding function $r : \mathbb{R} \mapsto \mathbb{R}$ is assumed with $r(x) = \lfloor x/\delta \rfloor + \beta$, δ the interval size and β some offset, although, w.l.o.g. it is assumed that $\beta = 0$ holds throughout this chapter. The discretized data $r \circ f = (r \circ f_1, \dots, r \circ f_n)$ is separated into connected components, so-called slabs, such that two adjacent

points $x, y \in \mathcal{S}$ are in the same slab if $\forall_{i=1}^n r \circ f_i(x) = r \circ f_i(y)$. Alternatively, x and y are in the same slab if there is a continuous path $p : [0, 1] \mapsto \mathcal{S}$ from x to y , i.e. $p(0) = x$, $p(1) = y$ such that $r \circ f(p(i))$ is constant for all $0 \leq i \leq 1$. Figure 6.3 illustrates the creation of slabs in a simple one-dimensional example with two functions.

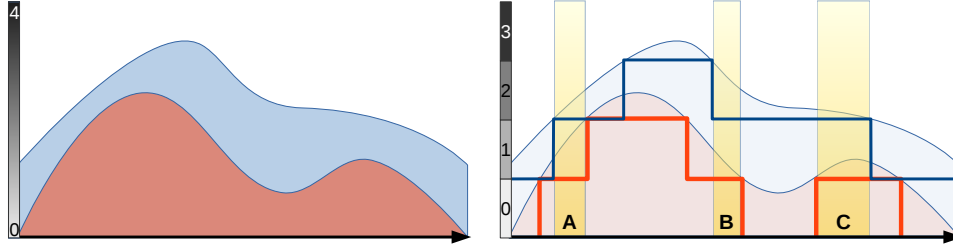


FIGURE 6.3: The left image shows two functions f and g colored in red and blue, respectively, with an image interval from zero to four. The right image overlays the functions with their discretized versions $\text{round}(f)$ and $\text{round}(g)$ colored similarly with $\text{round}(x) = \lfloor x \rfloor$, the floor function. The yellow colored regions A , B , and C indicate the slabs for the joint level set for $c = (1, 2)$.

To create the JCN, each slab is associated with a node. Two nodes are connected by an edge, if the corresponding slabs are adjacent. Note that in contrast to the reachability graph, the JCN uses undirected edges. A more detailed introduction on JCNs and its application can be found, for example, in work by Carr and Duke or by Nam et al. [29, 123].

Note how the size of the slabs depends on the *round*-function, such that the coarseness of that function allows a user to adjust the level of detail of the JCN. The changes in the JCN based on this coarseness parameter is another option to analyze the data. Note that this option does not exist in the Pareto set concept as presented in this thesis.

For an alternative approach to JCN, note that for the rounding function, $\text{round}^{-1}(c)$ is a continuous n -dimensional interval in \mathbb{R}^d with a lower bound l_i and an upper bound u_i in each of the n dimensions. Each slab containing x , with $f(x) = c$, can also be defined by a set of isosurfaces $f_i^{-1}(l_i)$ and $f_i^{-1}(u_i)$ for each $1 \leq i \leq n$. In Figure 6.3 the slabs A , B , and C are bounded by fibers with $f(x) = 1$ or $f(x) = 2$, and $g(x) = 2$ or $g(x) = 3$.

Note that the contour for $l_i \in \mathbb{R}$ is equivalent to the set of fibers for all $c \in \mathbb{R}^n$ with $c_i = l_i$ such that this set builds a *fiber surface* [34]. Hence, the domain \mathbb{M} can either be separated into slabs by a *round*-function or by a set equivalent to contours or fiber surfaces, respectively.

Increasing the number of used fiber surfaces to infinity without using the same contour twice or, analogously, an infinitely fine *round*-function will result in a structure equivalent to the Reeb space.

6.3.3 Translation into Directed Joint Contour Net

The JCN does not have a notation of ascending or descending direction between adjacent slabs. The definition of the Pareto set, however, is based on it. Hence, it

is unavoidable to integrate both concepts. Therefore, in this thesis the *directed Joint Contour Net* (dJCN) is introduced by replacing the undirected edges of the JCN with directed ones.

Definition 6.3.1. The directed Joint Contour Net $dJCN = (V, E)$ is a directed graph. Let $r : \mathbb{R} \mapsto \mathbb{R}$ be some rounding function such that the bijection b maps each connected component of $r \circ f$ to a node V . Given two adjacent components S_1, S_2 with points $s_1 \in S_1$ and $s_2 \in S_2$

$$((b(S_1), b(S_2)) \in E) \Leftrightarrow \forall_{i=0}^n r \circ f_i(s_1) \geq r \circ f_i(s_2).$$

With this, direction is introduced into the approximation of the Reeb space and this effect is used to compute an approximation of the Pareto Set. The structure of the dJCNs depends strongly on the rounding function. By refining that function, i.e. by decreasing the interval size and thus the size of the individual slabs, the number of slabs and therefore the number of nodes in the dJCN increases. Therefore, a series of rounding functions can be defined with the identity function as the limit, which leads to a series of dJCNs with the original domain as the limit. Such a series is called the *refinement series* of dJCN.

The limit of refinement

A Pareto maximum is a point x with a neighborhood $U(x)$ around x such that each point $y \in U(x)$ is either $x \prec y$ or $x \succeq y$. Alternatively, Pareto maxima can be defined through the global ascending set $C^+(x)$, such that x cannot reach any other point $y \in S$, i.e. no monotone increasing path p^+ to a dominating point y with $y \succeq x$ exists.

Analogously, Pareto minima are defined as points which cannot be reached via a monotone path from a dominated point y with $y \preceq x$. In both cases, y is not restricted to any neighborhood. To conclude, Pareto optimal points are surrounded by incomparable points, which makes them Pareto minimal and maximal simultaneously.

Obviously, this notion can be extended to slabs quite easily. A slab is called Pareto maximal if it cannot reach a dominating slab. On the other side, a slab is minimal if it cannot be reached from a dominated slab. For dJCNs, this simplifies to: A Pareto maximum slab has no outgoing edges and a minimum has no ingoing edges. As a reminder, in dJCNs there is only an edge if at least one of the rounded scalar functions changes.

The following set of examples illustrates how closely critical slabs and the Pareto set are related, even under this straight-forward definition. This is demonstrated by applying dJCN and Pareto set calculations to the LAMPS data set [30] and some artificial data sets. With this analysis, further relationships between both concepts are revealed.

LAMPS stands for “Limited Area Mesoscale Prediction System”. It is an atmospheric simulation, which is distributed with the VIS5D system, and provides a set of 10 scalar fields, including wind, temperature, pressure, and specific humidity. The application focuses on the east/west component (U) and the north/south component (V) of a simulated wind field.

Note that as a preprocessing step, the LAMPS data is smoothed with a Gaussian filter [125] before calculating the Pareto set, as described in Section 4.5. This removes smaller, noise-based features, which is automatically done by the *round*-function in the dJCN method. For more details see the corresponding section in Chapter 5.

In the following figures, the dJCN and the Pareto set for this and other, artificial, data sets are presented. In the right images, Pareto extrema are colored red, green, or yellow, depending on whether their ascending, descending, or both sets are empty, respectively. In the left images, the nodes of the dJCNs corresponding to critical slabs are colored red and blue, depending on the existence of only incoming or only outgoing edges, respectively. Otherwise, the nodes are colored green. The size of the nodes also indicates the size of the corresponding slabs. Note that in the later images, green nodes are reduced to transparent points to avoid obscuring.

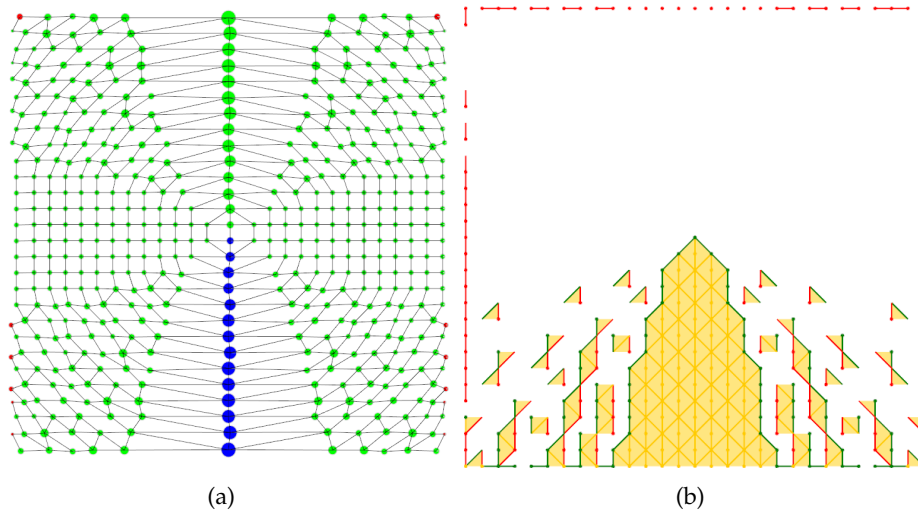


FIGURE 6.4: dJCN and Pareto set for a synthetic example based on a height and a distance function.

Figure 6.4 shows the dJCN and Pareto set for a synthetic example based on two fields. At each point, the first field is equal to the distance towards the center of the domain, while the second field is equal to the height in y -direction. Note how the pyramid-shaped component of Pareto extrema in the lower middle corresponds to the row of critical nodes at the same location in the dJCN image. This figure also shows some issues of the Pareto set calculation with degenerated functions, especially locations where adjacent vertices have the same function values. Such functions together with the triangulation of the data result in faulty Pareto extrema left and right to the mentioned pyramid-shape and the domain border. Note that this does not seem to be an issue for the calculation of the dJCN.

To consider further synthetic data sets, two and three Gaussian functions, respectively, are used to create a multifield (see Section 3.3). Therefore, for each individual field, a set of, fixed points in \mathbb{R}^2 for images 6.5(a) and 6.5(b) and in \mathbb{R}^3 for images 6.5(d) and 6.5(c) is given. The Gaussian function value is calculated as the sum of the exponentiated and weighted distance maps for these fixed points. For these examples, between the underlying functions only the location of the fixed points is moved by some degree.

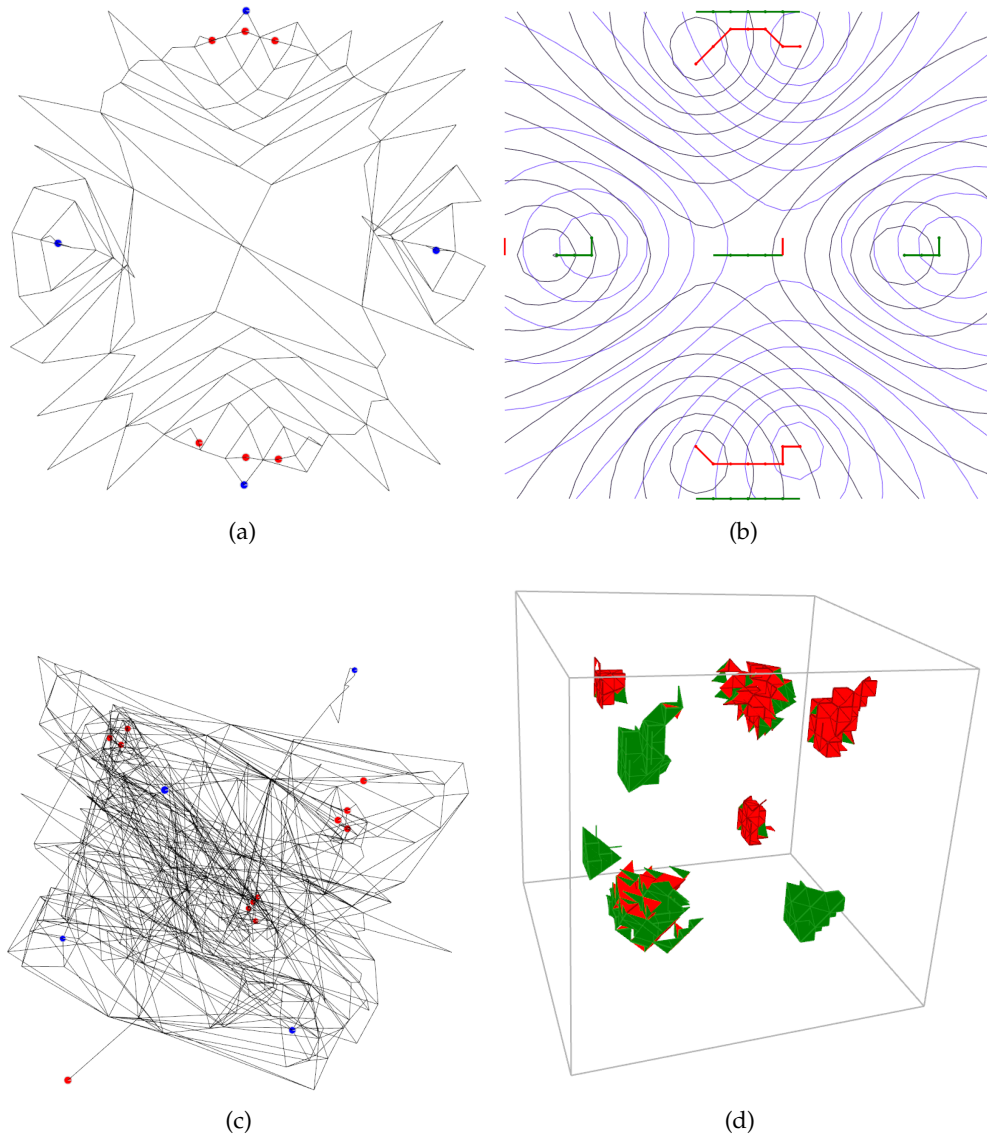


FIGURE 6.5: dJCNs and Pareto sets for two synthetic examples based on a set of Gaussian functions in 2D and 3D, respectively. Critical slabs in (a) and (b) and the Pareto extrema in (b) and (d) are colored similarly, with minima shown in red for both concepts, and maxima either in blue (dJCN) or green (Pareto sets). Also, (b) shows isosurfaces in blue and gray for the 2D functions.

As with the previous images, note the correspondence between most of the connected components of Pareto extrema with the groups of colored nodes in the dJCN. However, there are also Pareto extrema which cannot be linked to a critical slab. This is for example the case for the central component in Figure 6.5(a) and the larger two components in Figure 6.5(d), which are both bordered by triangles colored in red and green. Those components correspond to the location of saddle points in terms of a single-field topology. Therefore, and also due to large slab widths (see the isosurfaces in Figure 6.5(b)), the corresponding nodes in the dJCN have incoming as well as outgoing edges. Hence, 6.5(a) does not contain critical slabs in the image center, even though 6.5(a) clearly shows Pareto extrema at this position.

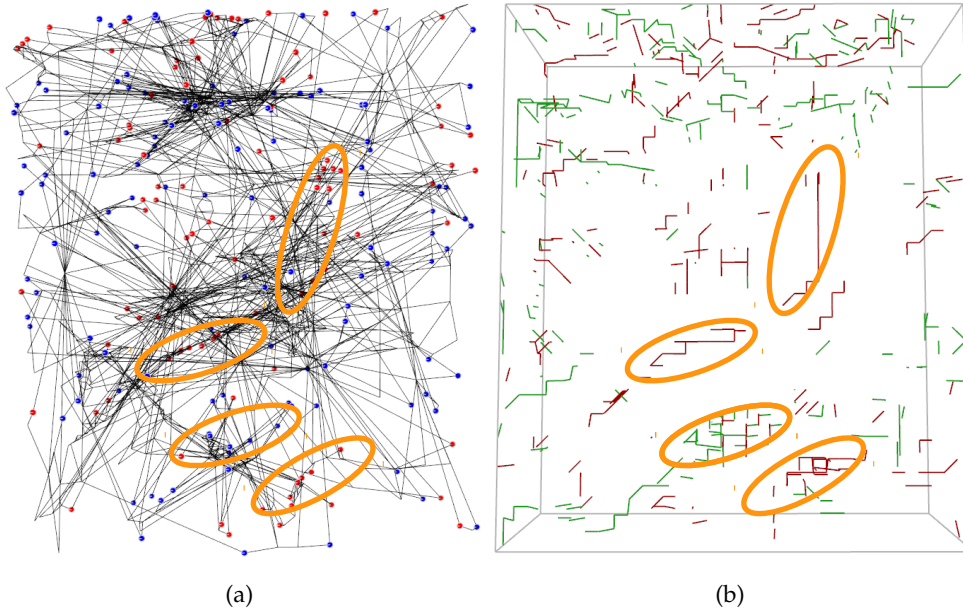


FIGURE 6.6: dJCN and Pareto set for the multifield function based on the U and V component of the LAMPS data set. The grid has over 21000 vertices which results in 95200 tetrahedrons. For the dJCN, the slab width is set to 5 for each underlying function. The calculation of 895 nodes and 1667 edges took around 25 seconds.

The last example is based on the LAMPS data set introduced in the beginning of this section. Figure 6.6 presents the Pareto set and the dJCN, for which each node is positioned in the barycenter of its corresponding slab. The figure shows how both visualizations have problems with critical elements in 3D data. Interactivity, i.e. rotation, zoom, etc., with the dJCN and the Pareto set helps to identify corresponding components of Pareto extrema and groups of critical nodes. While some of these pairs are highlighted through yellow circles a future, complete identification of all links has to be computer-based, for example based geometric distance between critical slabs and components in the Pareto sets.

Note that some critical nodes in 6.6(a) do not seem to correspond to any Pareto extremum in 6.6(b). However, also note that Pareto extrema that lie on the frontal domain border are made completely transparent to avoid occlusions.

A main similarity between Pareto sets and dJCN illustrated through these examples can be proven directly.

Theorem 6.3.1. *In the limit of a refinement series of dJCNs the union of all Pareto extremal slabs is equivalent to the Pareto set.*

Proof. The limit of the series of the rounding functions is the identity. Therefore the limit of the slabs are just single points or areas with the same scalar value and the applied definitions are equivalent. \square

6.3.4 Recognition of Pareto maxima and minima through a dJCN

In coarse refinements, connected components of Pareto maxima and minima can be merged into one slab. This raises the question at which step in the refinement series

all Pareto features are present in separated slabs in the dJCN. In order to decide when all connected components of Pareto maxima and minima are finally represented in their own critical slabs, the concept of persistence is applied.

Lemma 6.3.2. *After finite many refinement steps each Pareto maximal and minimal slab intersects exactly one connected component of the Pareto set.*

Proof. A merge of two connected components of Pareto maxima or minima into the same slab is only possible, if the refinement interval is larger than the persistence between any two points of the components for any of the scalar fields. Otherwise, the slabs containing them would be separated by at least one edge. Hence, if the interval size is smaller than the smallest persistence between any two components over all scalar fields, all connected components of the Pareto set are separated from each other.

Note that, after this step is reached and under the assumption of a monotone refinement series, further refinement cannot merge existing slabs and thus undo this separation. \square

Note that from this point on, for the remaining part of the chapter, results mainly from the master thesis by Jan Bormann [16] and from conference proceedings of the TopoInVis 2017 [17] are used.

6.3.5 Recognition of Pareto optima through a dJCN

Pareto optimal regions are maximal and minimal at the same time. By definition, no connections to adjacent slabs are allowed. Hence, at any slab border, one scalar function has to be on the upper and one on the lower bound of the rounding function. In practical cases, this is only true in the limit of the refinement series.

For example, assume a triangulation of a 2-manifold with three functions and three vertices: A, B and C. The values are $f(A) = (0.5, 0, 0)$, $f(B) = (0, 0.5, 0)$ and $f(C) = (0, 0, 1)$. Notice that all points in the triangle are incomparable to each other. Each function only increases in the direction of one vertex and decreases if it moves towards the other two. So, all points in the triangle are Pareto optimal. But, as exemplarily shown in Figure 6.7, for all rounding functions the slabs have connections and therefore do not fulfill the condition for Pareto optimal slabs.

The above examples show a general pattern. Scalar functions may divide the simplices at different points. In consequence, slabs will have connections although, they contain only incomparable points. This can be avoided by generalizing the notion of Pareto extremality.

6.3.6 Definition of ϵ -Pareto extrema

To summarize the problematic of the above definition of Pareto optimal slabs, let x , y , and z be three adjacent but incomparable points with $f_i(x) < f(y)_i < f(z)_i$ and $f_j(x) > f(y)_j > f(z)_j$. Assume furthermore, that a rounding function r exists such that $r \circ f_i(x) = r \circ f(y)_i = r \circ f(z)_i$ but also $r \circ f_i(x) < r \circ f(y)_i < r \circ f(z)_i$. Thus, all points are in separate slabs and connected with directed edges. Hence, none of these slabs are identified as Pareto optimal, because the definition only considers directly adjacent slabs.

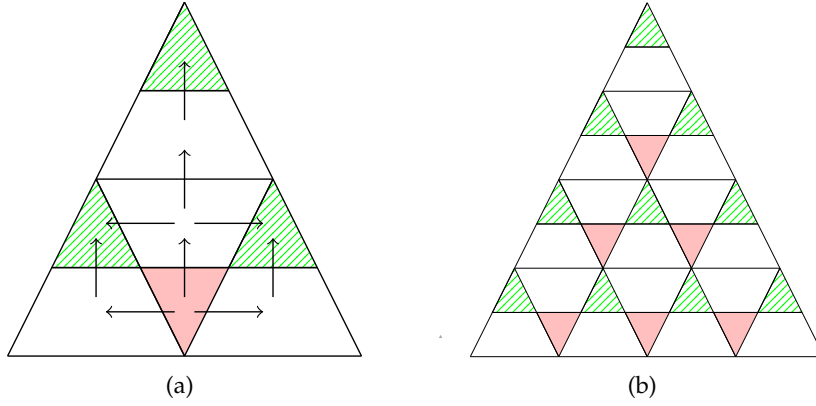


FIGURE 6.7: An example with Pareto optimal points in slabs which are not marked as such. Pareto maximal slabs are marked in green. Pareto minimal slabs are shown in red, while white areas correspond to regular slabs.

Therefore, the general idea is to define a neighborhood around each point in which the criteria for Pareto extremality are ignored, thus allowing it to consider longer paths in the dJCN beyond the direct adjacent nodes. The neighborhood is defined by the Chebyshev distance [28] $CDist_f(x, y) := \max_{i=1}^n |f_i(x) - f_i(y)|$ for any two points $x, y \in \mathcal{S}$. A point is called ϵ -Pareto maximum iff it cannot reach a point with Chebyshev distance of more than ϵ via a monotone increasing path. In other words,

Definition 6.3.2. Given an $\epsilon > 0$, a point $x \in \mathcal{S}$ is defined as a ϵ -Pareto maximum, iff

$$\forall_{y \in \mathcal{S}} x \rightsquigarrow y \Rightarrow CDist_f(x, y) \leq \epsilon.$$

Analogously, ϵ -Pareto minimum and optimum are defined and again, extend this notion to ϵ -Pareto extremal slabs. For completion, $\mathbb{P}_\epsilon(f)$ denotes the set of all ϵ -Pareto extrema, i.e. ϵ -Pareto maxima, minima, and optima. By definition, this notion is an over approximation of the Pareto set, and if $\epsilon = 0$ holds, it converts to the original definition.

Proof. To see this equality, ϵ is replaced in the above definition with zero such that $CDist_f(x, y) \leq 0$ has to hold, and therefore $f(x) = f(y)$. Assuming that $x \rightsquigarrow y$ is true, $f(x) = f(y)$ implies that all points along the corresponding path between x and y have the same multifield value. However, since f is assumed to be a smooth Morse function, this is only true for $y = x$. Hence, $\{y \in \mathcal{S} \mid x \rightsquigarrow y\} \setminus \{x\}$ has to be empty for x being a ϵ -Pareto maximum ($\epsilon = 0$). This is equivalent to the definition of Pareto maxima based on global ascending sets, see Definition 4.3.2 and Lemma 4.3.1. \square

As a reminder, the example in Figure 6.7 shows that the above definition of Pareto optimal slabs cannot identify Pareto optimal slabs as such although they contain Pareto optimal points, even for very fine rounding functions. The definition of ϵ -Pareto optima, however, is designed to provide just that. To prove this, additional notations are needed. Therefore, with the help of the gradients, the most increasing and most decreasing scalar field for each point is identified. For each Pareto optimum x and each direction v , the quotient of the absolute values of the most increasing and most decreasing gradient over all separate fields f_i is called the *changing rate*

of x towards v . The maximal changing rate is denoted with ch . Note that the changing rate is well defined, since x is assumed to be a Pareto optimum, i.e. surrounded by incomparable points such that at least one ascending and one descending field exists in each direction. Also, f is assumed piecewise linear, such that the gradients are piecewise constant. Hence, the maximum of all changing rates exists.

Lemma 6.3.3. *Let the interval size of the rounding function be smaller than the quotient of ϵ and ch , then all Pareto optimal points are in ϵ -Pareto optimal slabs.*

Proof. Given a Pareto optimal point $x \in \mathcal{S}$, w.l.o.g. $x = 0$, some path p starting at x in direction v , the slab containing x can only be classified as ϵ -Pareto optimum, if along the path p at least one field increases and one field decreases, both with a value change of more than δ , the interval size of r . Then, the corresponding path in the dJCN can neither be classified as ascending nor descending path. Furthermore, these value changes need to appear within a Chebyshev distance of ϵ with respect to f around x . Otherwise, the ascending or descending subpath of p within the Chebyshev distance is sufficient to make x neither a ϵ -Pareto maximum nor ϵ -Pareto minimum, respectively.

Hence, it is possible to approximate, from below, the number increasing and decreasing value changes along p within Chebyshev distance of ϵ that would result in slabs changes along the corresponding path in the dJCN. If both numbers are larger than 1, the p cannot be used to disqualify x as a ϵ -Pareto optimum.

Therefore, let y be the first point along p for which $\max_i(|f_i(x) - f_i(y)|) \geq \epsilon$. W.l.o.g. it can be assumed that the maximal argument of this Chebyshev distance is $i = 1$ such that $f_1(x) - f_1(y) = \epsilon$. For all other arguments, i.e. $i \neq 1$, it holds that $f_i(x) - f_i(y) < \epsilon$. Furthermore, since x is Pareto optimal, it is possible to find a minimal point y' between x and y and an argument j , w.l.o.g. $j = 2$, with $f_2(x) - f_2(y') < 0$. Otherwise, an ascending path from x to y exists and therefore x could neither be Pareto maximal and, thus nor Pareto optimal.

The number of slab changes c_1 based on f_1 is therefore approximated from below: $\frac{f_1(x) - f_1(y)}{\delta} = \frac{\epsilon}{\delta}$. For an approximation of c_2 , the slab changes based on f_2 , it can be assumed that both f_1 and f_2 have the flattest ascend and descend possible and, w.l.o.g. have $f_1(x) = f_2(x) = 0$. In other words, $f_1(t) = a \cdot t$ with $\frac{f_1(y)}{y}$ and $f_2(t) = b \cdot t$ with $\frac{f_2(y')}{y'}$.

Following this assumption, the location of y is defined through ϵ and the gradient of f_1 :

$$f_1(y) = \epsilon = a \cdot y \Rightarrow y = \frac{\epsilon}{a}.$$

The number of slab changes c_2 based on f_2 is therefore:

$$\left| \frac{f_2(y)}{\delta} \right| = c_2 \Rightarrow c_2 = \left| \frac{\epsilon \cdot b}{a \cdot \delta} \right| = \epsilon \left| \frac{b}{a} \right| \cdot \frac{1}{\delta}.$$

Since it is assumed that $i = 1$ is the maximal argument for a point y with Chebyshev distance $CDist_f(x, y) = \epsilon$, it follows that $|f_1(x) - f_1(y)| \geq |f_2(x) - f_2(y)|$ and thus, $\left| \frac{b}{a} \right| \leq 1$. Therefore, $c_2 \leq c_1$ is the lower boundary for direction v .

This argument, illustrated in Figure 6.8, can be done analogously for every path originating from x to conclude the proof. □

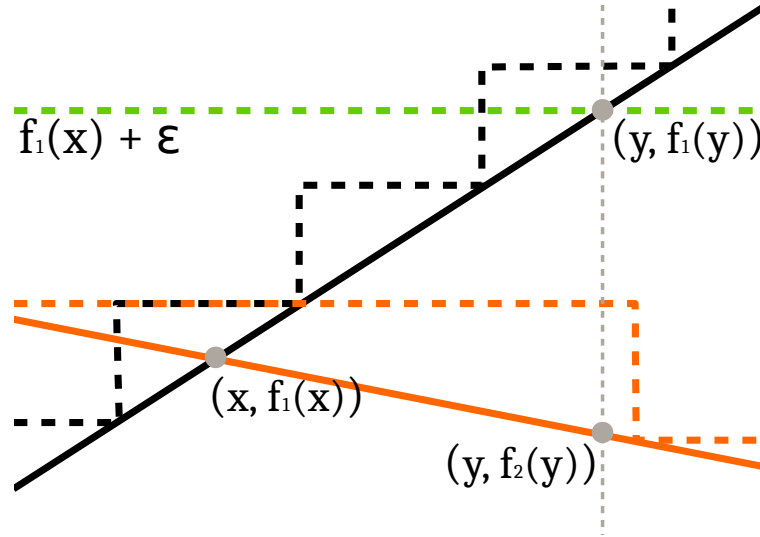


FIGURE 6.8: Illustration of the number of slab changes between point x and y for two fields f_1, f_2 .

With a lower bound for the interval size given, the following theorem is a direct result.

Theorem 6.3.4. *For almost all refinement steps all connected components of Pareto extrema are contained in ϵ -Pareto extremal slabs.*

Note that by decreasing ϵ to zero the over approximation is reduced, but more slabs to compute. On the other hand, by choosing a rounding accuracy a limit of the approximation rate of Pareto slabs can be set.

6.3.7 Proof-of-Concept Examples over constant 2D-grids

In the first results in Figures 6.4, 6.5, and 6.6, the dJCN is computed with the naive definition of critical slabs through modified code provided by Haemish Carr, while the following visualizations result from Jan Bormann's implementations.

Regarding the latter, to support the changing ϵ and interval size, Bormann uses an alternative implementation of the slabs in contrast to existing literature [29, 123]. Assuming the grid is evenly triangulated, each triangle is recursively separated into equally sized subtriangles based on the interval size and the corresponding rounding function r . Therefore, triangles are split, if for any two centroids x and y of the resulting subtriangles and any field f_i , $r(f_i(x)) \neq r(f_i(y))$ holds. In other words, the triangulation is refined until the the slab borders are closer to the triangle edges than to the triangle centroids.

In a second step, the slabs are approximated by iteratively clustering triangles. Therefore, two neighboring triangles are grouped together, if their centroids are inside the rounding interval, i.e. $r(f_i(x)) = r(f_i(y))$, for all fields i . This allows for fast

recalculation of the slabs if the interval size is changed.

An easy example showing all types of ϵ -Pareto extrema is based on three shifted Gaussian functions, similar to the two used for the first images in Figure 6.5.

Figure 6.12 shows how the connected components of the Pareto set are over-approximated by ϵ -Pareto extremal slabs. Note how the shape of the extremal slabs in 6.9(a) is similar to the Pareto extrema in image 6.9(b). Also note the classification into Pareto minimal, maximal, and optimal slabs that matches the classification of the Pareto extrema in 6.9(b). This effect can be improved with increased refinement at the cost of a higher number of slabs. The implementation is not restricted by the number of fields or the complexity of the data, as demonstrated in the example shown in Figure 6.10.

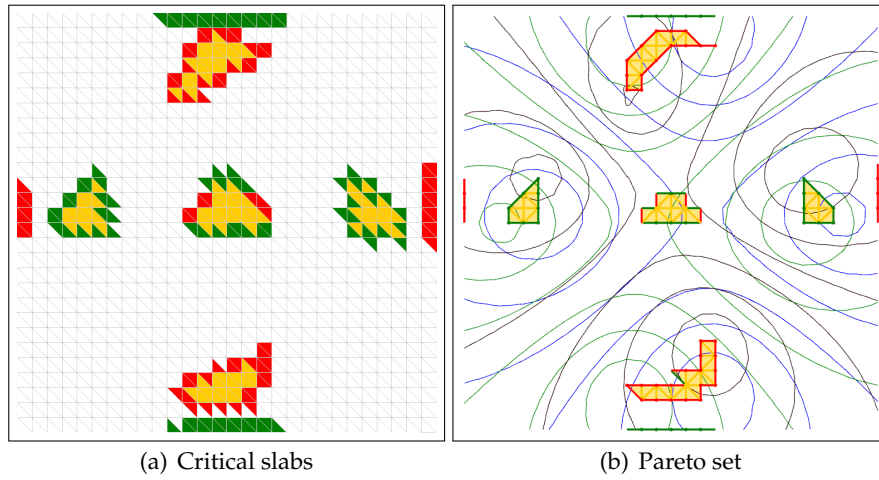


FIGURE 6.9: Pareto extremal points and slabs colored in green, red, and yellow depending on their type. In (b) the three individual fields are indicated by a selection of contour lines.

For the fields in Figure 6.10, again Gaussian functions with slightly moved extrema are used. Movement occurs parallel to the horizontal and vertical axes such that the components of the Pareto set result in cubic shapes. As in previous sections, some irregularities (marked with blue circles) can be noticed. They can be attributed to the coarseness of the triangulation. Note how those irregularities appear in the critical slabs 6.10(a), depending on the interval size of the rounding function.

Furthermore, currently neither dJCNs nor any other approach can indicate where critical slabs appear after further refinement of the rounding function. This prohibits stepwise, local refinement, i.e. refinement only around critical slabs.

Reachability Graph

To visualize additional information about the Pareto set, a connected component in the Pareto set *reaches* another, if a monotone ascending path between any two points of those components exist. Obviously, this also implies that a path between the slabs containing those points exist in the dJCN. Furthermore, since the ϵ -Pareto

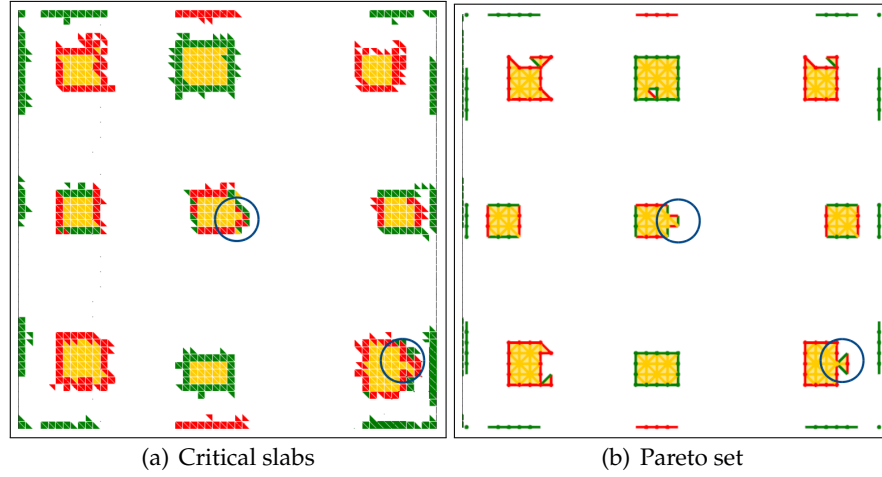


FIGURE 6.10: Pareto extremal points and slabs colored as in Figure 6.9 but the data set contains an increased number of fields and each individual field has more single-field extrema.

optimality is also defined through reachability, the same ideas are applied to prove similar properties for the reachability graph, especially regarding its approximation by a dJCN.

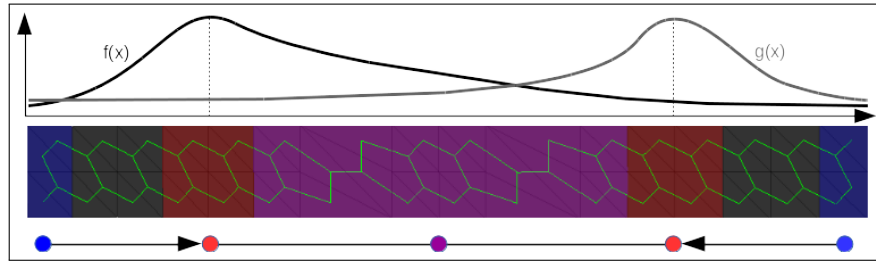


FIGURE 6.11: Two 1D fields (top), a dJCN (center) for a 2D extension of the fields, and the corresponding reachability graph (bottom).

The next example shows two 1D scalar fields $f_i : [0, 10] \mapsto \mathbb{R}$ in the upper part of Figure 6.11, each with a distinct, obvious maximum in the center. The fields are extended to 2D with $f_i(x, y) = f(x)$ for $y \in [0, 1]$ and $i \in \{0, 1\}$. The grid is triangulated, and the resulting slabs are shown in the center part of the figure using Bormann's implementation. Colors indicate the slab status, while green lines connect neighboring triangles only if their centroids are comparable, i.e. if they are either dominating or dominated.

A Union-Find algorithm [158] combines neighboring Pareto extremal slabs of the same type and provides the nodes for the reachability graph. Edges are created by running a path-finding algorithm on the dJCN. In this case, breadth-first-search is used. In summary, this results in the following procedure:

- Create the dJCN
- Calculate the ϵ -Pareto set

- Identify clusters of Pareto extremal slabs through Union-Find
- Build the reachability graph through Breadth-First-Search on the dJCN

6.3.8 Discussion

In this section, the connection between JCNs and Pareto sets are not only visually shown, but their relation is also proven and quantified. In detail, the following contributions to the study of multifield topology-based analysis techniques are provided:

- The first, naive definition of critical slabs shows potential for visualization, but is disproved to be a direct translation of the complete Pareto set.
- Therefore, the ϵ -based definition of Pareto sets is introduced, both in context of simplicial complexes and the dJCN.
- The presented theorem proves, in the limit of the refinement series and under the ϵ parameter, that critical slabs are equivalent to Pareto sets such that the dJCN can be used as an approximation of the Pareto set.
- Furthermore, through the intermediate definition of ϵ -Pareto minima and maxima, the error between Pareto sets and dJCNs and the resolution of the rounding interval can be limited.

In summary, it is possible to approximate the Pareto set through the dJCN and to limit its error to some degree. The dJCN, furthermore be used to approximate the reachability graph using simple path finding algorithms for graph structures in contrast to existing, computationally expensive approaches on piecewise linear structure. It is also shown that the error limit between the critical slabs and the Pareto set is reached after a finite amount of refinement steps. However, it is not possible to efficiently calculate ch and thus the necessary resolution of the rounding function beforehand. Note that a naive calculation requires roughly the same computational effort as the calculation of the Pareto set itself, thus providing no faster runtime.

The results are implemented and applied to a set of artificial data sets with multiple fields and discussed the implications of this section for the general relations between Pareto sets and the Reeb space. While the latter could provide an error-less identification of the Pareto set, so far, it is apparent that a fast method to construct the reachability graph out of the Reeb space exist. With this, directions for future work include utilizing simplification approaches based on JCN or the Reeb space to efficiently remove local structures from multifield visualizations.

Reeb space

The JCN is an approximation of the Reeb space, a generalization of the Reeb graph in multifield data. In detail, given the equivalent class $x \sim y$ iff both points belong to the same path connected component of the preimage $f^{-1}(f(x)) = f^{-1}(f(y))$, the Reeb space is the space of all equivalent classes with the quotient topology inherited from \mathcal{S} (see Section 3.2.3). Obviously, JCN slabs converge to these preimage components for an infinitely fine rounding function. Thus, it is reasonable to discuss the Reeb space as a means to calculate the Pareto sets and the reachability graph.

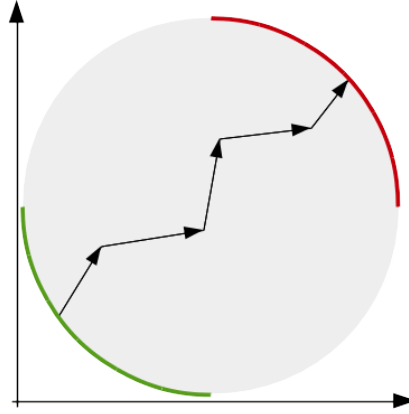


FIGURE 6.12: Pareto extremal points and slabs colored in green, red, and yellow depending on their type.

The Reeb space has the advantage of being continuous and not an approximation, as the JCN. Thus, the original definition of Pareto sets can be applied instead of the detour of ϵ -Pareto set. Let for two points $x, y \in \mathcal{S}$, $x \rightarrow y$ hold, such that a continuous ascending path from x to y exists. Then, there is also an ascending path in the Reeb space from the equivalent class containing x to the class containing y . Therefore, corresponding to the definitions in Section 6.3.3, points in the the Reeb space can be defined as Pareto minimal, Pareto maximal, or Pareto optimal. Like Jacobi structures, the Pareto set can be extracted from the Reeb space. See for example the work by Chattopadhyay et al. [36] for comparison. The extraction is improved by the fact that the Pareto set can be considered a subset of the border of the Reeb space projected onto the image space of the multifield.

However, the main advantage of the dJCN, i.e. the graph structure allowing for a quick computation of an approximated reachability graph, is not given anymore. It is therefore reasonable to hypothesize that the calculation of the ascending and descending paths is equally expensive in both the domain \mathcal{S} and in the Reeb space.

Chapter 7

Application Scenarios

Besides defining the Pareto sets and showing their relationship to existing techniques, it is important to demonstrate their potential application in real-world scenarios to better understand, when and how Pareto sets might be a useful tool for visualization and exploration. This also provides an overview of possible extensions of the Pareto set concept and its combination with other visualization tools. Following this goal, the application of Pareto sets in three scenarios that can broadly be labeled as problems requiring multifield data analysis are described in this chapter. In each of these scenarios, the data sets are either based on simulation ensembles, contain multiple variables or time steps, or are otherwise semantically related. Individual fields are associated with the same triangulated domain, given either in 2D or in 3D. Hence, the definitions and algorithms from Chapter 4 can be applied.

7.1 Vortex Detection

One possible application of the multifield method is its usage to integrate multiple criteria for vortex detection in fluid flow or weather simulations. The general idea is that, while many criteria for vortex detection exist, each of them incorporates an individual level of result uncertainty. This is due to different criteria models, uncertainty or missing data in the initial conditions of a simulation, or other factors. It is therefore reasonable to consider a variety of different criteria at once to compensate for this uncertainty.

Using the Pareto set for this combination is described in the following two subsections, whereas the first describes a scenario from fluid simulation, while the second considers a similar scenario from a meteorological simulation. Note that both subsections are basic, proof of concept scenarios for the usability of Pareto sets. Other, more detailed scenarios are introduced in Section 7.2 and 7.3.

7.1.1 Scenario 1: Vortex detection in fluid simulations

This scenario is based on the *Can data set*, which is already mentioned in Section 5.4. The images and results can also be found in earlier work [91]. The data itself results from a CFD (computational fluid dynamics [190, 176]) simulation of flow in a fluid-filled cylinder (courteously provided by M. Rütten, DLR Göttingen). It is represented on a tetrahedral mesh of approximately 750K elements and describes the transient flow of a highly viscous fluid in a cylinder at rest. In this simulation, fluid motion is induced by the rotating top of the cylinder. The data set is a simple example of a flow simulation with the aim to identify vortices using several vortex criteria. In the focus of this subsection are the Q -criterion as well as λ_2 , vorticity, and pressure values and the corresponding extremal points. Such simulations are often

found in aircraft or car design, where the localization of air turbulence and vortices are essential for a fuel efficient design.

As a scalar quantity, the vorticity ω reflects the local strength of rotation and becomes maximal at the vortices' centers but can also indicate shear in boundary layers, similar to the fluid pressure criteria. As another vortex indicator, the Q-criterion q is positive where rotation dominates shear and is also maximal in vortex centers. The variable λ_2 (see Jeong and Hussain [97] for definition) becomes minimal in vortices, therefore, to get a good match with the maxima for the other fields, the negation of this field, $-\lambda_2$, is considered instead. See, for example, Jeong and Hussain [97] for further details on those criteria.

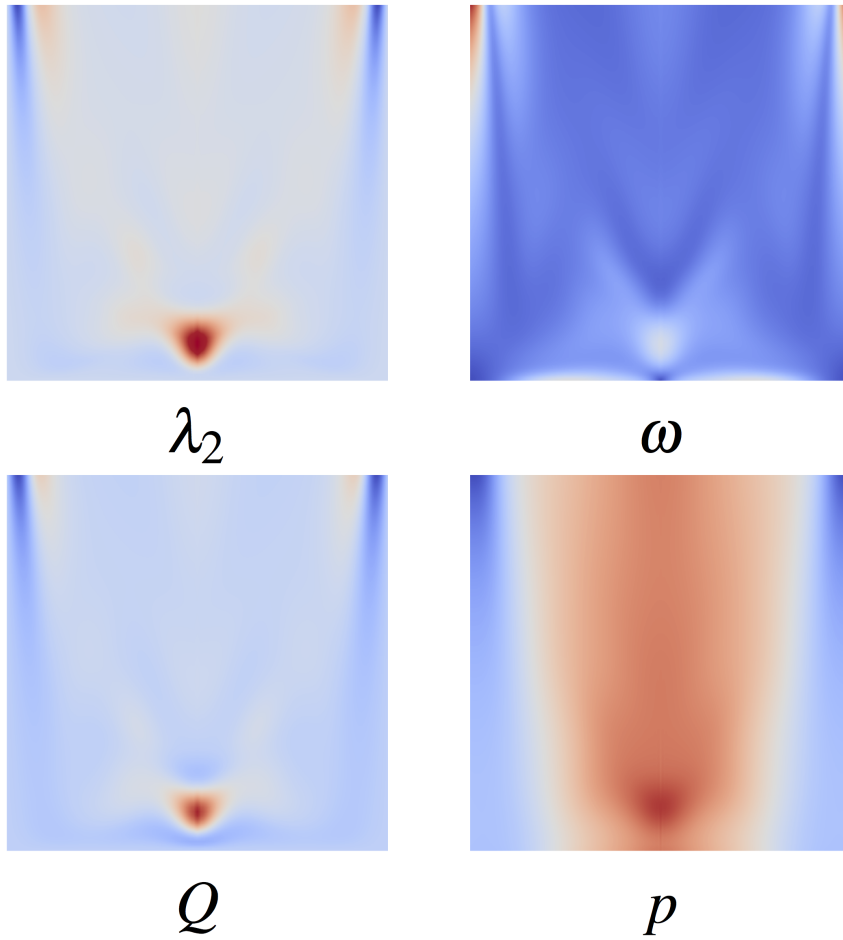


FIGURE 7.1: Color maps of the scalar fields going from blue for low values to red for high scalar values.

In summary, all criteria indicate vortices via extremality, however disagree on the exact location of detected vortices and might contain mis-classifications. Therefore, the Pareto set seems like a natural choice to integrate these measures for a more reliable detection, since it provides extremal regions, elements of $\mathbb{E}(f)$, on which all criteria agree.

Next to the 3D data, additional input is generated as a two-dimensional slice parallel to and containing the cylinder's central axis. As the flow is relatively smooth

and exhibits a high degree of rotational symmetry, this is sufficient to identify meaningful structures. The images in Figure 7.1 illustrate the component scalar fields on the slice.

The Pareto set for the combination of the four fields in Figure 7.1 is shown in the two images in Figure 7.2. While a common maximum in the center bottom is obvious in the color maps of the individual fields (Figure 7.1), other features, for example the three layers of wing-like structures in the center of the domain along the central axis, are almost impossible to identify without the Pareto set. Images 7.2(b) shows a color-based domain decomposition as all ascending and descending sets of all connected components of the Pareto set are visualized in different, but transparent colors. Thereby, the presented colors are the results of overlapping ascending and descending sets and their corresponding transparent colors. Hence, the domain is visually separated by color and each color corresponds to an equivalence class of points that can reach and can be reached by the same set of connected component of the Pareto set. This visualization is more time efficient than the actual calculation of the equivalence classes but open to color bleeding. However, it is sufficient and hints at the complexity of creating the reachability graph. The decomposition furthermore shows how the mentioned wing-like structures extend and are connected to each other.

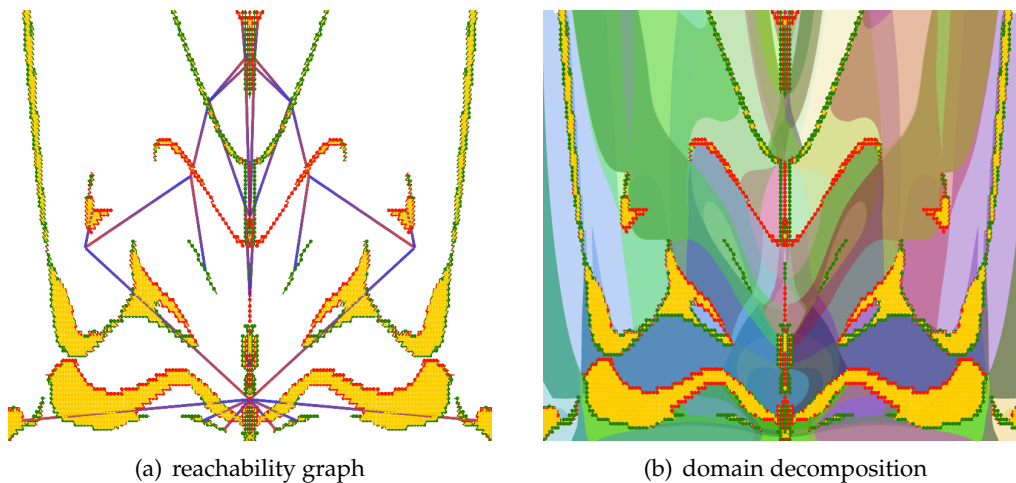


FIGURE 7.2: The Pareto set for the fields in Figure 7.1. (a) shows also the corresponding reachability graph, while (b) shows a visualization of all ascending and descending sets of all connected components contained in the Pareto set.

In general, note how all criteria agree to the extent that vortices appear either along the central axis or around the bottom or sides of the container, but not in between. The latter is indicated by the absence of Pareto extrema in this area. Also note that the Pareto optimal region at the bottom is rather large, meaning that area where the four fields are incomparable, i.e. disagree in a common direction/behavior, is extensive.

Other visualizations are shown in Figure 7.3, where the Pareto set contains only two fields, λ_2 and the Q-criterion. Image 7.3(a) provides a 3D view on the can data

set. Instead of the wing structures seen in the 2D slices, it is now possible to identify layers of cone-shaped structures around the central axis. Figure 7.3(b) shows the Pareto extrema of an axial cross-section of the data set augmented by $H^\pm(x)$. The cones are colored in light green and red, respectively. The images demonstrate one possible option to explore the data, by highlighting the ascending and descending set of a regular point to indicate a vortex region up to the point where the vortex breaks down. This approach shows close details but no global behavior. This might become slightly confusing due to the many, small ascending and descending cones. Therefore, in Figure 7.3(c), $H^\pm(x)$ are colored gray while the incomparable points are left white.

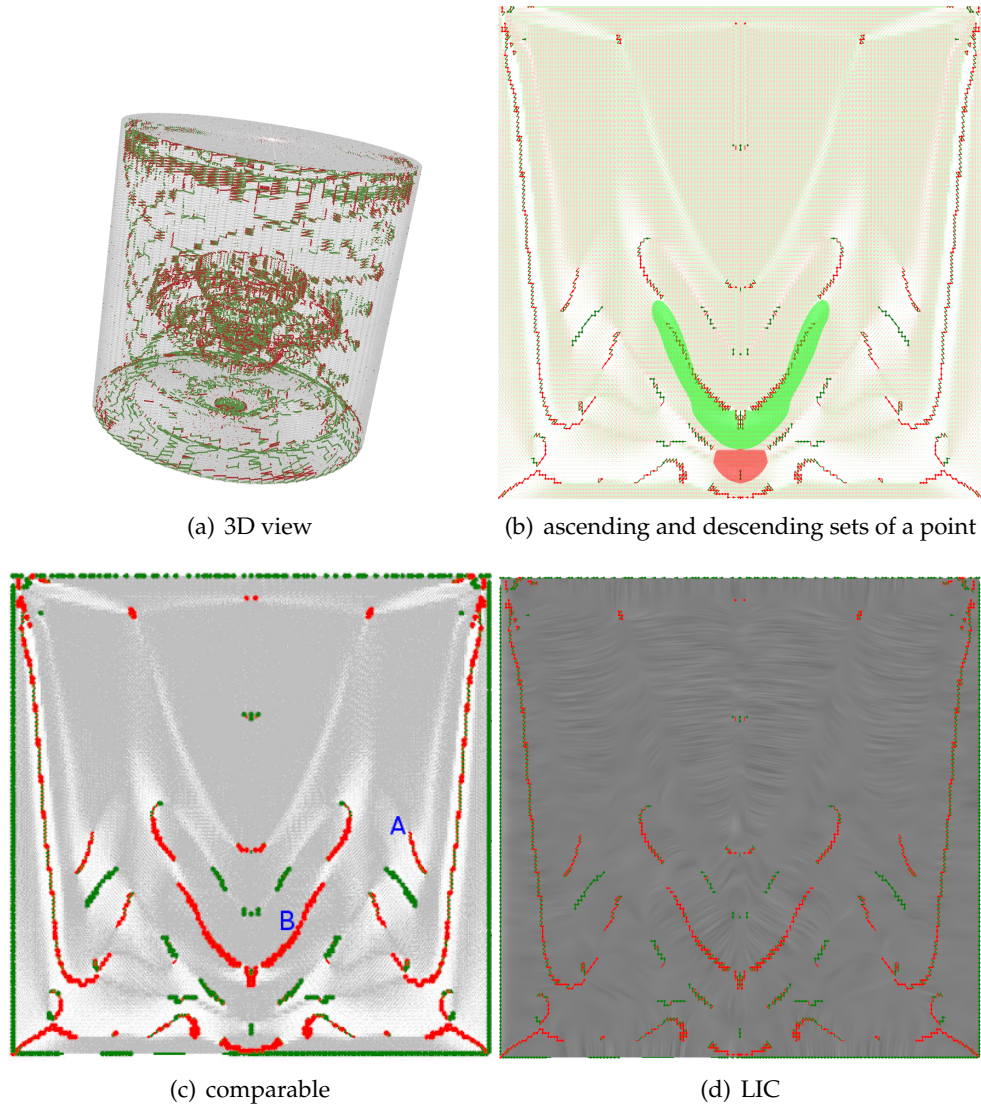


FIGURE 7.3: Pareto extrema in a can flow simulation. While (a) depicts Pareto extrema in the 3D volume, the other images focus on an axial slice of the dataset. The different visualizations highlight different aspects. See main text for details.

Note how the Pareto region A is surrounded by narrow, almost invisible cones, while the cones around region B are wider. This provides a rough estimation on how

much the gradients of λ_2 and the Q-criterion match in this Pareto-optimal region. Therefore, while A will consist more of incomparable points and less of true optima, i.e. optima in at least one of the criteria, B is more like a hillside for both fields. For this thesis, a possible approach is investigated using the size of $H^\pm(x)$ inside the triangle to measure the agreement between the individual fields in the following Section 7.3.

Alternatively, Figure 7.3(d) shows a line integral convolution (LIC) visualization, using the angle bisections of $H_\sigma^+(x)$ as vector with the size of $H_\sigma^+(x)$ as length. Note that in regions with large ascending cones, i.e. where the individual fields agree with each other, the image gets fuzzier. In contrast, in regions with smaller cones, i.e. the ascending and descending paths are more narrow, the image stays sharp. This provides a very rough and approximated view on the local ascending and descending sets, the averaged gradients of the individual fields, and some indication about their agreement. While this is still not an optimal approach to visualize the multifield, it provides a good overview of possible extensions to the Pareto set visualization as well as a proof-of-concept.

7.1.2 Scenario 2: Atmospheric Hurricane Data

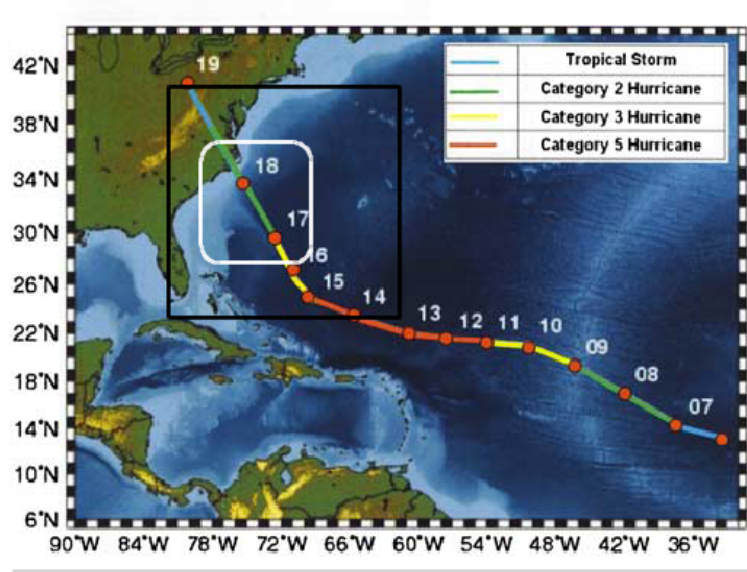


FIGURE 7.4: The path of hurricane Isabel during September 06-19 2003. The black box indicates the spatial area and the white box indicates the time window contained in the data set. The background image containing the hurricane track and the development of its intensity is from a paper by Gautam et al. [66].

In climatology and meteorology, simulation data sets play a key role in analyzing atmospheric processes and events. Such data sets commonly consist of time-varying, multivariate 2D and 3D fields - properties that aggravate their visualization and analysis. Using the IEEE Visualization 2004 Contest Hurricane Isabel data produced by the Weather Research and Forecast (WRF) model (courtesy of NCAR [124] and the U.S. National Science Foundation (NSF)), the Pareto set is used to identify vortices, corresponding to features of the Hurricane Isabel. It demonstrated how Pareto Sets

can be utilized to identify and visualize global features that correspond to common extremal areas in a set of spatially corresponding scalar fields. As in the previous subsection, the multifield can consist of a combination of different vortex criteria, but here, different height levels are considered instead of just one axis-parallel slice. Additionally, arrays of subsequent time steps are studied to track changing behavior of selected variables over time as well. The visual result were also presented during the 32nd Conference on Environmental Information Processing Technologies by Kathrin Hüb [92].

Spatial Dimensions:	Longitude 83.00W - 62.00W (2139.00 km) Latitude 23.70N - 41.70N (2004.00 km) Height 00.00 km - 19.84 km
Spatial Resolution:	500 x 500 x 100 grid cells ($\delta x \approx 4.28$ km, $\delta y \approx 4$ km, $\delta z \approx 0.198$ km)
Number of Time Steps:	48 ($\delta t = 1$ h, Sept 16-18, 2003)
Number of Variables:	13

TABLE 7.1: Hurricane Isabel data produced by the Weather Research and Forecast (WRF) model, courtesy of NCAR and the U.S. National Science Foundation (NSF) [106].

The hurricane Isabel moved 2003 along the US coast as illustrated in Figure 7.4. The image shows the hurricane's path, the observed spatial domain (black box) and the observed time window (white box) contained in the data set. For more details about this illustration, see Gautam et al. [66]. In all images for this subsection, a satellite images of the corresponding observed spatial domain (North Atlantic Ocean west of Florida retrieved from Bing [12]) is used as background to provide geographical context to the different Pareto sets that are given as overlays.



FIGURE 7.5: Features of hurricane Isabel. The image is from the International Space Station (15. September 2003) [180]).

Note that this application is not part of the IEEE Contest but a presentation for the 32nd Conference on Environmental Information Processing Technologies (EIPT) which took place during the 96th Annual Meeting of the American Meteorological Society 2016 (AMS) in New Orleans, USA. Hence, some publications regarding the visualizations and analysis of the hurricane data set exist [1, 117, 151, 66, 47, 187, 46, 13, 70]. However, these works only use single- or multifield tools without considering topological aspects. The data is provided in context of the IEEE Visualization 2004 contest [106] and is summarized in the Table 7.1.

While possible, it is not reasonable to calculate the Pareto set of all $48 \times 13 = 624$ fields at once. The semantic relation between fields from different time steps and different variables is simply too low to create meaningful results. Hence, subsets of fields are chosen in which at least two of the three criteria time step, variable and height layer, are the same. The application goal is inspired by work from Joshi et al. [151] that analyzes and visualizes hurricane features like the hurricane eye or the eye wall, shown in Figure 7.5 and the location and time of start of the hurricane dissipation.

Time-varying near-ground pressure

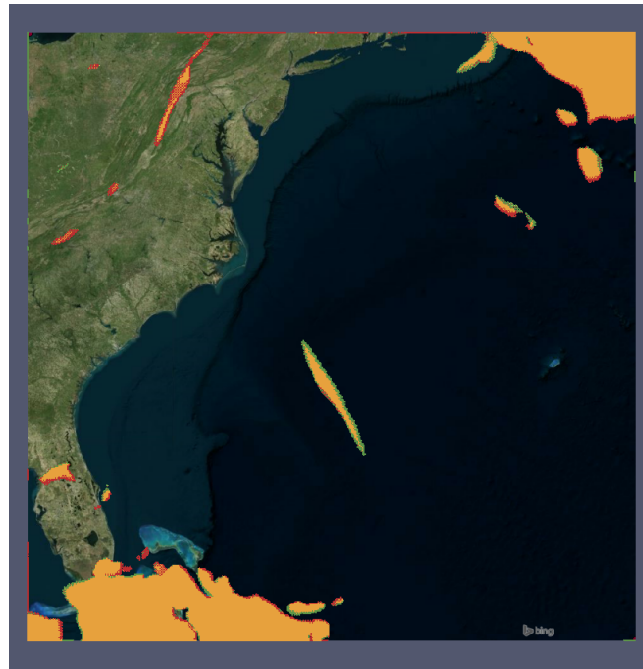


FIGURE 7.6: Pareto set, colored in green, red, and yellow depending in the Pareto extremal type, for air pressure simulated 20 meter above sea level. The multifield contains 21 fields, each representing one hour time steps (see Figure 7.8 and 7.7).

Figure 7.6 shows the Pareto set for time steps 10 to 30 for the pressure near the ground level (approximately 20m above sea level). Each time step corresponds to one hour such that similar behavior of the individual fields can be assumed. In the focus is a high pressure center in the image center which is marked by a Pareto optimal region (yellow area) surrounded by Pareto maxima (green line). It clearly indicates the general direction of the hurricane center. This statement is verified in

Figure 7.8 and 7.7. The figures show the air pressure for first time step ($t = 10$) and, in a multiview, for the remaining ten time steps, respectively, as gray scale map with the Pareto set from Figure 7.6 as an overlay. Dark gray indicates high air pressure such that each image shows a clear pressure maximum moving along the hurricane path shown in Figure 7.1. Note that a multiview of all individual fields, as in Figure 7.7, is not suitable to present all pressure maps at once such that either the enlarged view of selected images (see Figure 7.8) or a summary, for example with the Pareto set concept, is necessary.

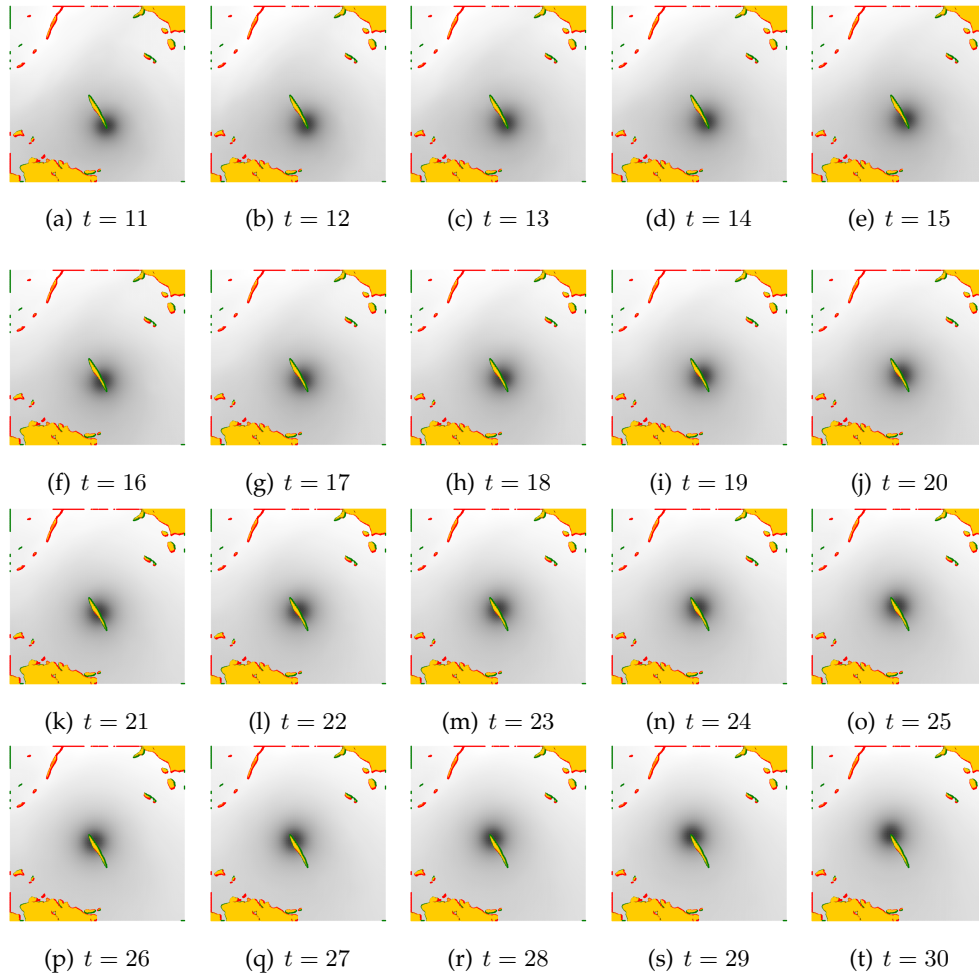


FIGURE 7.7: Pareto set as shown in Figure 7.6 with a grayscale map based on the air pressure 20m above sea level at different time steps t indicated by the image labels.

While the Pareto set provides a good summary of the 20 time steps and compensates for uncertainty in the individual fields, it does not indicate if the hurricane moves towards or away from the coast line. However, note that it is possible to move a time window, for example 10 steps, along the time line to compensate for this drawback and allow the trace of the hurricane center over time.

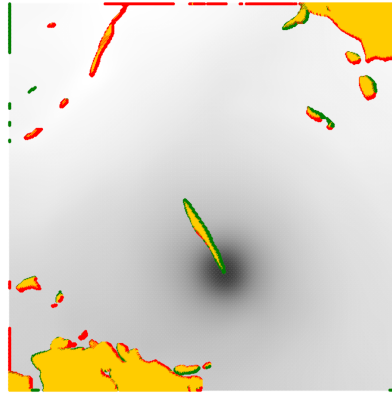


FIGURE 7.8: Pareto set as shown in Figure 7.6 with a grayscale map based on the air pressure 20m above sea level at time step 10.

Height-varying wind speed

The images in Figures 7.9 provide the wind speed at time step 20 for different sets of height layers. Especially with 3 to 6 layers, two main characteristics of the hurricane, the eye and the eye wall, are clearly visible as Pareto minima (green) and Pareto maxima (red), respectively. It shows that the eye is very stable throughout the height of the hurricane, resulting in a small point-like component in the center. Note on the other hand, that due to the rotation of the hurricane, a circular Pareto maximal component corresponding to the eye wall appear. There, local maxima in the individual fields or different height layers are positioned in a spiral pattern and thus cumulate to almost a circle in the Pareto set.

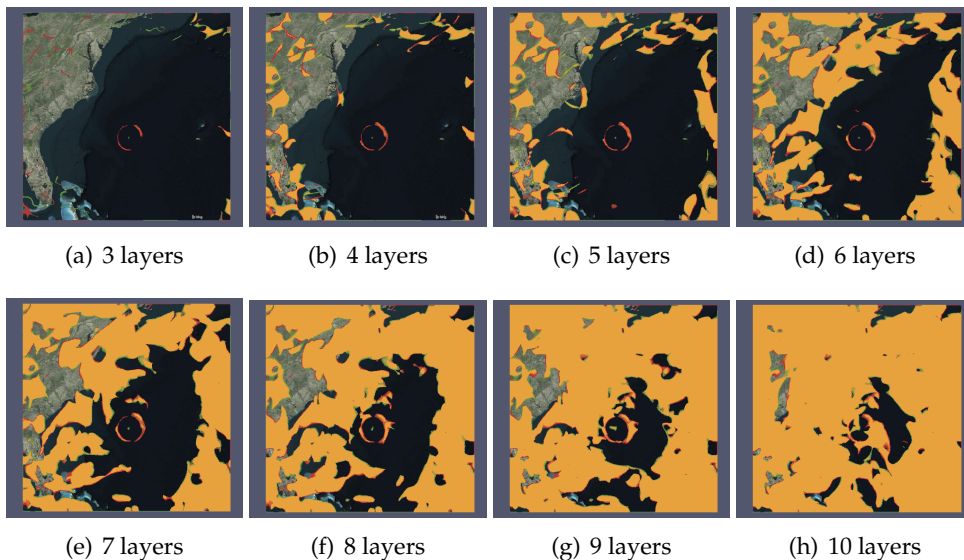


FIGURE 7.9: The Pareto set for different numbers of hight layers at time step 20, i.e. around September 16, 20:00 UTC.

As a third remark, note that a hurricane starts to spread out in the higher atmosphere such that the conical eye wall becomes wider in higher levels and also that the wind speed at sea level is almost incomparable to the wind speed above 15km. Hence, the Pareto-optimal regions become too large to provide meaningful results if too many height layers are included in the multifield, as shown in image containing 10 height layers.

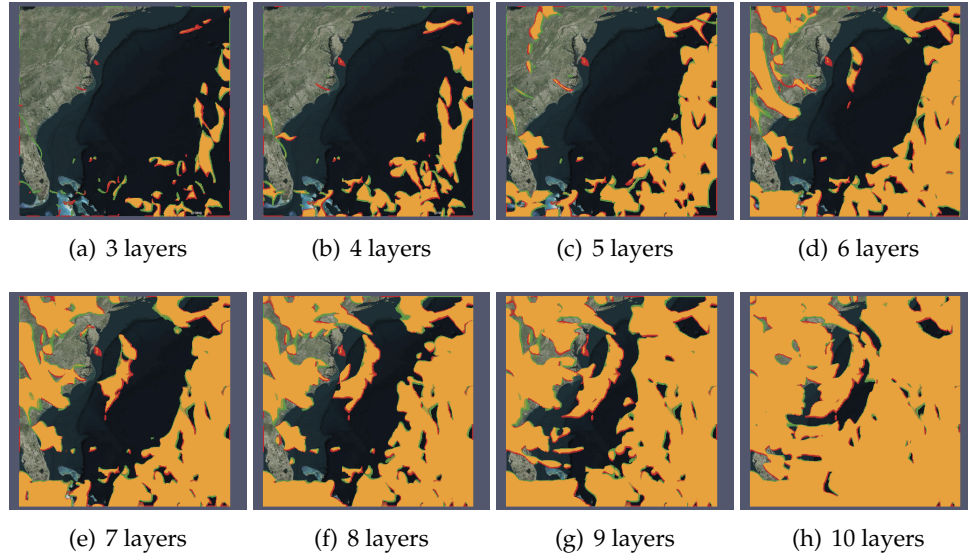


FIGURE 7.10: The Pareto set for different numbers of height layers at time step 45, i.e. around September 17, 21:00 UTC.

The same height layers of wind speed can be analyzed for a different time step, for example for the last one. This is shown in the images in Figure 7.10. Due to occlusion of the satellite image in the background, the presumed location of the hurricane is marked with a blue circle. Again, features like the hurricane eye and the eye wall can be identified. However, it appears as if size, shape and distance between the features identified by the Pareto set method changed, when comparing the image of this later time step to that of the earlier. Those changes might be founded in the hurricane slowing down to a category 2 hurricane (see Figure 7.4).

Due to the lower category and therefore less rotation, the wind speed maxima in the eye wall move less when adding higher layers and therefore resulting in less circular Pareto maximal regions. The center, less vertical, seems to tip over such that the corresponding Pareto minimal region is more a line than a point. And furthermore, the height level at which the hurricane starts to fan out is lower, such that the Pareto optimal regions appear larger in Figure 7.10 as opposed to those with the same set of height layers in Figure 7.9.

In summary, the analysis of Pareto sets not only allows to locate hurricane features, but also, when compared over time, can indicate when the hurricane starts to dissipates. Quantifiable measures like the component size corresponding to the hurricane eye or its distance to the components corresponding to the eye wall, can even provide a measurement for the degeneration.

7.1.3 Results

Throughout this section, it is shown that the Pareto set is able to detect vortices, even with distinct features like hurricane eyes and corresponding eye walls. However, the actual values of the individual fields are not visualized, since the presented concept is only based on the gradients. While this is intended by the Pareto set design, it prohibits explicit visualizations showing which field is responsible for Pareto optimality inside the Pareto set.

The scenarios also show how reliant the visualization is on the semantic relation between the fields. Even a small set of fields, as shown in Figure 7.2 or 7.10, can cause large Pareto optimal regions, i.e. areas of incomparable points, even when the fields are supposed to represent the same phenomenon, for example the position of vortices. The results of this section support the previous assumption that Pareto sets are more likely to be useful for ensemble data set, in which the fields are based on similar simulation models or measurement techniques but, for example, with changed initial conditions or parameters.

7.2 Quality Control in Car Manufacturing

The second application is based around real-world data from car manufacturing processes, in detail their quality control. Quality control and maintenance is a vital part of the car manufacturing process, since it is mandatory that the produced parts fit together in each assembly step. Hence, divergences between parts and their respective target shape must be uncovered, such that the responsible assembly step or machine can be further evaluated or repaired.

Nearly every car part has to be checked to identify incorrect working stations or devices and/or wear of the machine parts. For the quality control, the manufactured part is clamped into a frame or held by a robot arm, resembling spatial positions as precisely as possible for a good registration of the actual part and its quota. Small deviations are accounted for by the measurement software. For sheet metal, there are several methods of quality control. In this section two techniques are in the focus: tactile measurements and optical measurements.

Tactile measurements are conducted, e.g., by a high precision robot arm that leads a small probe head to the sheet surface, generating one measurement point for each feature in a predefined inspection plan [111]. For optical measurements, a scanning device, e.g. a stereoscopic camera system, is used to generate point data for a contiguous part of the surface, creating millions of measurement points regardless of the tactile measurement plan [135]. Both methods can be used independently or together. However, new possibilities arise for visualization and analysis from the denser and orders-of-magnitude larger set of data points resulting from the optical system.

To investigate these possibilities, the Pareto set approach is tested using so-collected optical measurements. These data sets are represented as triangulated point clouds (meshes) with millions of triangles, given for example as STL-files [153]. For practical applications, it is mandatory that the measurements are always connected to the correct quota, measurement plan and measurement plan version. To facilitate this task, most companies use a measurement data management system (MDM).

The method proposed in this scenario can, for example, be included into the widely used eMMA software suite by Kronion [104]. That way, it can easily be integrated into the workflow, providing a useful complement during all steps from prototyping to series production.

The results and work, presented in the following can be found in previous work [90] as well as in projects by Nils Feige in cooperation with Kronion [104].

7.2.1 Scenario Overview: Data Set and Application Goal

The Pareto set visualization is applied to a set of seven 3D-scans of a trunk lid. The first scan is arbitrarily chosen as quota with a grid consisting of over 1.7 million triangles. The distance maps from the quota geometry to the other six scans are depicted in Figure 7.11. Note, that since the scans are all of the same trunk lid, the results visualize problematic areas for the 3D scanner rather than actual wear of the machines involved in building these lids.

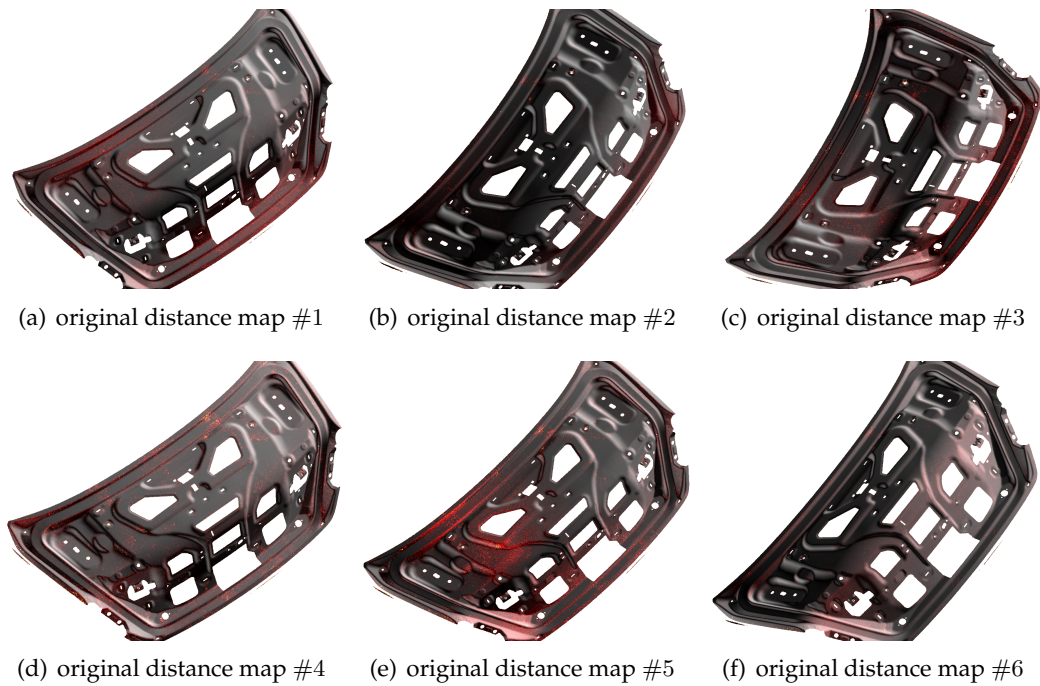


FIGURE 7.11: Selection of distance maps based on 3D scans of the same car trunk lid, displayed color-coded on the quota grid. The color scale ranges from black (zero deviation from the quota) over red to yellow (high deviation from the quota).

Nevertheless, the general goal is to enhance the insight into the data by aggregating information from an entire set of measurements into one simple view that automatically highlights areas where all distance maps show the same behavior. The resulting visualization should indicate the location of areas with systematic errors in the production process, which in general might stem from worn punching press tools or other sources like faulty transport processes. These locations should be furthermore distinguishable from other locations that are generated through uncertainty caused by the sensor arrays [21]. Thus, the method, in contrast to simply

averaging the distance values over multiple measurements, should not show random outliers in one function as a systematic error.

Preprocessing of the measurement data

To calculate the distance map between the quota and an actual (produced) part, two triangulated meshes are needed. While the actual data always stems from a triangulated 3D point scan, the quota could also be a triangulated CAD geometry. The distance between every scan point of the actual part to the surface of the triangulation of the quota is then saved in the closest vertex of the quota triangulation. If multiple points of the actual part are mapped onto the same vertex, the distance value is averaged. Hence, every scanned actual part results in a distance map over a common grid. Given multiple car frames of the same production line, their scans result in a scalar multifield such that each vertex stores a vector, each entry representing a distance between the quota and one scans.

For some applications, it is not necessary or not feasible to scan the entire surface of an actual part. Thus, only partial scans are produced and provided for comparison. In these cases, it should be avoided to consider regions on the quota lacking corresponding data from the actual part as an error in the actual part. Vertices on the quota without an equivalent on the actual part are assigned zero distance. Hence, such distance mapping is highly asymmetric.

Initial inspection of the data revealed it to be strongly noisy. Large areas of all distance maps had only small distance values with little, local changes. It is hypothesized that these small-scale variations are due to small errors in the scanning process. To filter out this noise, a Gaussian smoothing method is applied to each distance map. This smoothing assigns each vertex a new distance value based on the weighted average of its neighbors. The neighbors are identified by traversing along the edges using breadth-first search, only considering neighbors within a fixed depth level. The weights are Gaussian functions based on the Euclidean distance between the neighbors and the vertex [125].

7.2.2 Definition of the Weighted Pareto Sets

To implement the visualization, the Pareto set is utilized. As a reminder, note that every maximum of any field f_i is also a Pareto maximum of f and is therefore included in one of those connected components $\mathbb{E}(f)$. But, as presented in the 1D example in Figure 7.12, a Pareto extremal region does not indicate that all fields f_i have an extremum inside this component.

Figure 7.12 presents the Pareto maxima, Pareto minima, and the Pareto optima, marked by a red bar, a green bar, and a yellow overlay, respectively, for three 1D-fields f_1 , f_2 , and f_3 . (see Chapter 4). Note how only f_2 contributes to the Pareto extremal region labeled A , i.e. by omitting f_2 from the multifield, all points in A would become regular.

To tackle the application goal, the focus lies on regions in which most distance maps have a maximum, i.e. to which many fields contribute. Other Pareto optimal regions however, to which only a small number of fields contribute, are of less interest since those possibly count towards uncertainty in the optical scans. Hence, each

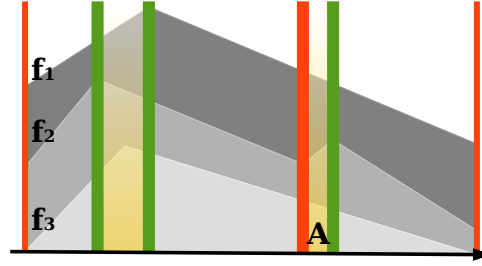


FIGURE 7.12: A simple 1D example with three functions f_1 , f_2 and f_3 and the corresponding Pareto extrema colored in red for Pareto minima, green for Pareto maxima, and (transparent) yellow for Pareto optima.

Pareto optimal region receives a weight corresponding to the number of contributing fields, to allow an automatic distinction between important and unimportant regions.

Definition 7.2.1. The weighted Pareto set $\mathbb{P}_w(f)$ for a multifield $f = (f_1, \dots, f_n)$ is a Pareto set with weighted components $\mathbb{E}(f)$. The weight function $w_+ : \mathbb{E} \mapsto \mathbb{N}$ is defined for all connected components $E \in \mathbb{E}(f)$ such that $0 \leq w_+(E) \leq k$ is the number of fields f_i for which at least one point $x \in E$ is maximal in terms of single-field topology, i.e.

$$w_+(E) = |\{x \in E \mid \frac{\partial}{\partial x} f_i(x) = 0 \wedge \frac{\partial^2}{\partial x^2} f_i(x) < 0\}|.$$

Given the distance map, in which high distance indicates high wear, only maxima need to be considered instead of counting all extremal points. A component $E \in \mathbb{E}$ can be measured as a region in which $w_+(E)$ of the individual fields have at least one maximum. High values of $w_+(E)$ indicate that inside this region many fields differ from the quota geometry and therefore advocate that systematic errors, for example by wear, appear in the production process.

In summary, while the Gaussian smoothing reduces noise and those maxima in the separate distance maps with only a small local value difference are eliminated, the weighting of the components in the Pareto set can remove those, which are only based on a small number of distance maps. Hence, a user can focus on those areas where a large number of distance maps indicate the location of relatively large distance, which are areas that might be affected by wear.

7.2.3 Results

The following figures show the weighted Pareto sets for those six fields, presented in Figure 7.11, after a Gaussian smoothing. Each caption indicates the number of smoothing iterations (I), the full width at half maximum (FWHM) that defines the width of the Gaussian bell curve, and the maximal depth level of the neighborhood (ND). See Section 5.1 for more details regarding the Gaussian smoothing.

In Figures 7.13, 7.14, and 7.15, the color scale encodes the following properties: Red indicates a regular vertex and a weighting of zero while the color scale from blue to green indicates Pareto extremal vertices with a weighting from one to six, meaning that one to six distance maps have a maximum in this area. Higher weights

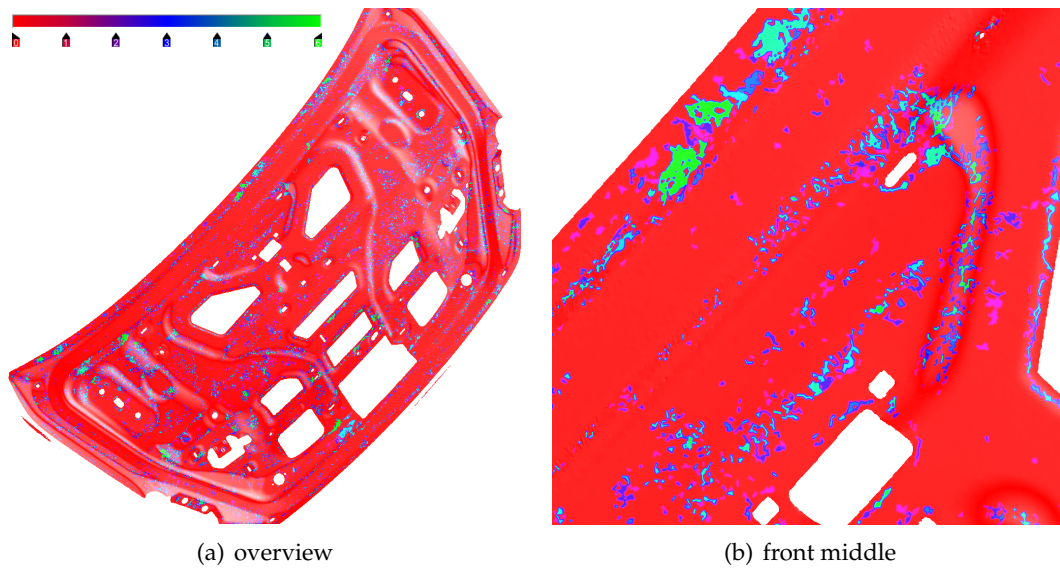


FIGURE 7.13: The weighted Pareto set on the trunk lid for the six distance maps shown in Figure 7.11 after a Gaussian smoothing with parameters I : 10, FWHM: 6, and ND: 5. Regions are colored based on their weights, going from red (0) over blue (3) to green (6).

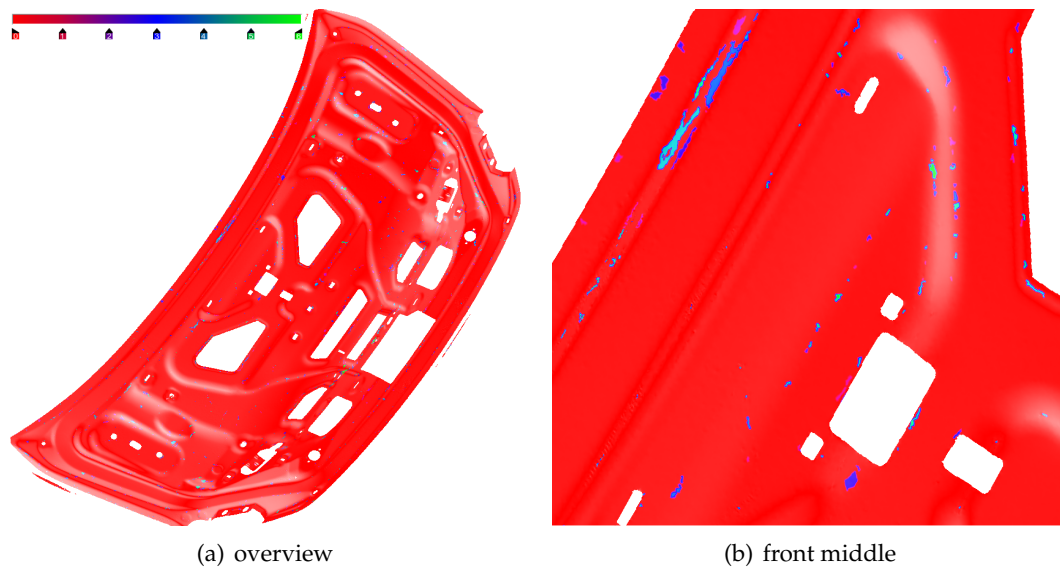


FIGURE 7.14: Image analogous to Figure 7.13 with different smoothing parameters (I : 20, FWHM: 3, and ND: 5). Note how the oversimplification removed most of the features visible in Figure 7.13.

indicate that there is an area of systematic errors, the detection of which is the main goal of the proposed algorithm. The color for the triangles and edges are linearly interpolated based on their adjacent vertices.

Figure 7.14 illustrates that the search for suitable smoothing parameters is necessary. While still some areas of interest exist (colored in green), most regions disappear due to oversimplification.

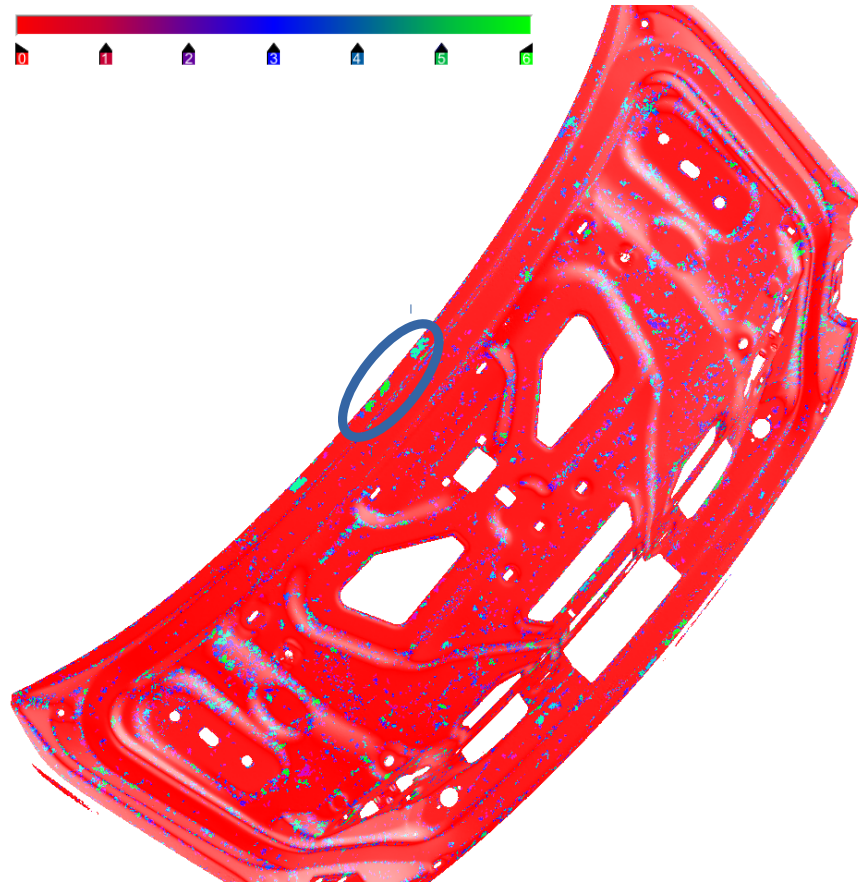


FIGURE 7.15: Image similar to Figure 7.13 (smoothing parameters I: 10, FWHM: 3, ND: 3). Note how most of the Pareto extrema can be found on slopes of the trunk lid.

As can be seen in Figures 7.13, 7.14, and 7.15, the most problematic regions, indicated by the green color, are those located on the slopes of the lid. Since the scanned part did not change during sampling, it is assumed that these are regions where the scanner produces the highest errors. While this result is understandable and supports the verification of the approach, other regions of interest can also be noticed, which could not easily be detected through observation of the separate distance maps. Note, for example, the encircled Pareto maximal area in Figure 7.15, which which cannot be explained easily. The high weights indicated by their light green color and the fact that the areas remain even under different smoothing parameters, see 7.13(b) and 7.14(b), suggest interesting behavior. All areas lead to focused analysis and improvement in understanding the data.

To further demonstrate the influence of the smoothing parameters on the resulting Pareto set the Appendix A include a set of images based on the same distance functions as Figures 7.13 and 7.14. The parameter setting for each image can be retrieved from the included parameter table.

In conclusion, the weighted Pareto set allows for a better distinction between different connected components in the Pareto set and thus providing more information about the individual fields and how they contribute to the Pareto optimality. Note that in this scenario, it is not necessary to know which distance maps are actually indicate wear. Instead, the focus in this application is rather laid on retrieving a quantifiable measure that implies systematic errors.

With respect to future work, since wear is a gradual process, averaging over multiple measurements before calculating the Pareto sets might be a promising approach in order to aggregate data stemming from similar wear-off stages. It is also possible that the Pareto regions move over the surface. In this case, feature tracking algorithms could be applied to highlight the temporal development of wear-off processes. Both can be combined into the time-window idea from the previous application scenario described in Section 7.1.

To further support a user-driven exploration of the data, other approaches are possible, including the representation of the Pareto sets through hierarchical abstraction. Here, connected components can, for example, be clustered into meaningful groups. Currently, the recalculation of the Pareto sets with new parameters is too slow for an immediate visualization of changes, which makes the performance improvement of the Pareto set computation a mandatory task in the near future.

7.3 Comparing Ensemble Members of Climate Projections

The third application scenario focuses on climate research and its challenges regarding visualization. In climate research, ensemble simulations are often carried out to capture the statistical behaviour of the climate system; the spread of the results represents the range of potential future climate behavior. Furthermore, it was shown that, within multi-model ensembles, the so-called "mean model" consistently outperforms individual models in nearly every respect [69]. Ensemble simulations can be used to assess uncertainties in the model-based prediction of atmospheric variables [45, 159]. These uncertainties are usually quantified using statistical measures applied to the varying results of the members belonging to an entire ensemble of climate simulations [40]. For this, the inter-model standard deviation (ISD) is a simple, but frequently applied summarizing estimate (see, e.g., [168]).

Inter-model standard deviation for a multifield f with at least two individual fields ($n > 1$) over a manifold \mathbb{M} is a map $g : \mathbb{M} \mapsto \mathbb{R}$, defined as

$$g(x) := \sqrt{\frac{\sum_{i=1}^n (f_i(x) - f(\bar{x}))^2}{n-1}}.$$

The term $f(\bar{x})$ is the mean value over all individual fields at point x , i.e.

$$f(\bar{x}) := \frac{1}{n} \cdot \sum_{i=1}^n f_i(x).$$

The ISD or square root of the variance measures how far the individual fields are spread out and thus quantifies the amount of point-wise variation or dispersion in the multifield [14]. Critical points in the ISD-map, especially maxima, can be used

to locate regions with high uncertainty in the climate projections. However, the ISD does not provide more information, neither locally, for example why uncertainty or dispersion exists at these locations, nor globally.

In the visualization community, the problem of analyzing climate ensemble simulations has been tackled using various approaches, ranging from tailored mapping techniques for a certain combination of key measures [19] to interactive frameworks for the multi-level exploration of ensembles [134].

Regarding visual analysis of climate ensembles, Poco et al. systematically define domain-specific intents for analyzing model similarity and develop a system aimed at exploring the differences between members in a climate ensemble [129]. In a further paper, Poco and co-authors propose an iterative approach to understanding differences between ensemble members using not one but several difference measures [130] and leverage this to understand the relation between criteria to and outputs of climate models. Similarly, Wang et al. [171] describe a visual analysis approach to examine parameter-dependence of climate model output. They utilize various linked views such as e.g. heat maps, and in particular develop a nested variant of parallel coordinate plots. Rautenhaus et al. [137] used a similar approach but also included 3D-visualizations like isosurfaces and normal curves in their linked views.

More general, considering multiple linked views, Potter et al. [133] conclude that systems employing these typically results in clearer presentation and improved visual analysis. While their approach is general in nature, they demonstrate this on weather forecast and climate ensembles, where this technique is particularly effective, and apply statistical visualizations specifically as individual views [132]. Their approach is captured in the ViSUS-CDAT system [134]. Focusing on predictive quality of climate ensembles, Böttinger et al. [19] systematically visualize variables of climate ensembles concurrently with predictive skill and ensemble spread.

In the application targeted in the current section, the Pareto set concept is not used as a stand-alone tool for the analysis of the data set under consideration (compare Section 7.1 and 7.2). It is rather combined with other visualization methods, such as simple color maps of statistical measures and glyphs.

The results presented in this section are created in cooperation with Kathrin Feige and Michael Böttinger and results are partially presented during the Vis in Practice Workshop [89] at the VIS Conference 2017, Phoenix, USA.

7.3.1 Scenario Overview: Application Goal

In the application scenario described in this section, the goal is not only to identify regions of similar and dissimilar behaviors of the individual fields in the ensemble data set, but also to enhance the data exploration. This includes visual feedback about the quality of similarity, thus allowing a more detailed differentiation between locations where the climate models behave more or less similar. Therefore, in addition to the previously applied binary separation of the domain into regular and Pareto extremal region, the goal is to build a continuous map between these two types. This map can then be combined with the visualization of other measures, such as the ISD. This is a valuable addition, since it enables a comparison between the neighborhood-based Pareto set approach and features appearing in the

ISD-map [191, 40], which is computed point-wise.

To furthermore enhance the exploration, it is important to visualize which fields contribute to the presented Pareto set. In the focus are the identification of outliers and subgroups of fields behaving similarly. This delivers deeper insight on how different climate models, initial conditions, or boundary conditions contribute to differences in the behavior of the resulting climate projections.

7.3.2 Enhanced Visualization of the Pareto Set

In the first examples in Section 4.1, the ascending and descending sets are visualized directly, colored in less saturated green and red, respectively. In contrast to that, the Pareto optimal regions are highlighted using (opaque) yellow regions, as is also shown in Figure 4.9. To target the application goal of showing more information than the binary domain separation into regular and Pareto extremal points, the direct visualization of the ascending and descending set is replaced using a grayscale map that corresponds to the *opening angles*.

As a reminder, the opening angle $\alpha^+(x)$ of a point $x \in \mathbb{S}$ is defined as the opening angle of the ascending cone $H^+(x)$ at x (see Definition 5.2.1), primarily in the case of \mathbb{S} being in a 2D simplicial complex with piecewise linear fields. For the semantic of α^+ , note that, if $\alpha^+(x)$ is small, at least two fields strongly disagree in a common ascending direction. This means that even small changes in the values belonging to these two fields can transform the regular point into a Pareto extremum. In contrast, if $\alpha^+(x)$ is large, all fields strongly agree and the area of common ascending or descending directions is large. Note that, due to the piecewise linear fields, α^+ is a piecewise constant functions of \mathbb{S} with values between 0 and π . Due to the linearity, the opening angle cannot be larger than π .

For further exploration, it is advantageous to delineate regions within the domain that show similar opening angles. Thus, intuitively speaking and with regard to the application focused here, the domain is decomposed into regions showing similar topological relationships between climate ensemble members.

Definition 7.3.1. *The discretized version of α^+ is defined as $\beta_\sigma^+(x) := \lceil \alpha_\sigma^+(x)/c \rceil$ denoted as decomposing function with some refinement value $c \in [1, \pi]$.*

The connected components of the level sets of β_σ^+ , i.e. $\{x \in \mathbb{S} \mid \beta_\sigma^+(x) = v\}$ for all $v \in \beta_\sigma^+(\mathbb{S})$, correspond to regions that are similarly close to becoming Pareto extremal. In the context of the incremental algorithm, level sets only have to be recalculated for triangles with changing defining indices and changing β^+ , if new fields are added to the analysis.

In this section, β^+ is visualized as a gray scale map over the domain, going from white for $\beta^+(x) = 0$, to dark gray for $\beta^+(x) = 1$, excluding those regions already colored as Pareto set. An example is shown in Figure 7.16, where the dark gray region becomes smaller, the more similar the fields under comparison are. Both images present two Gaussian functions, all with a maximum near the center. In Image 7.16(a) the maxima are further away than in Image 7.16(b). Thus, in the second image the fields are more similar than in the first one which is reflected in the lighter gray colors in the background (a grayscale map based on β^+).

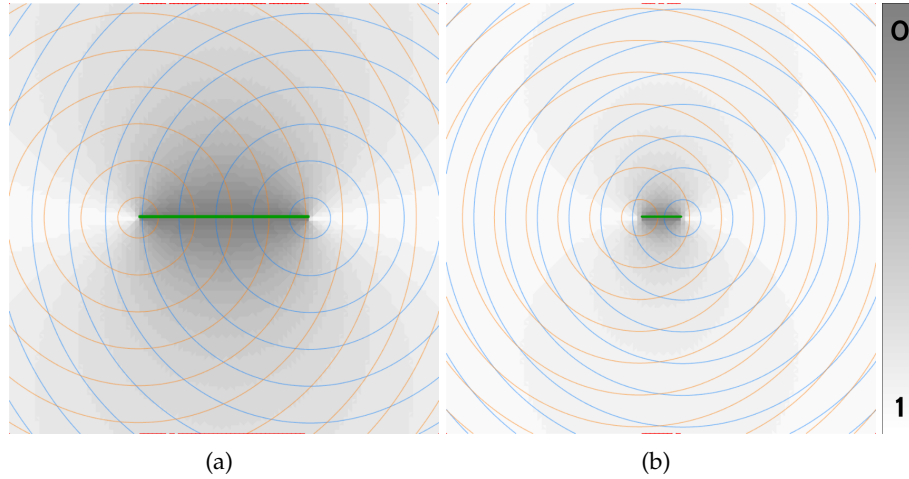


FIGURE 7.16: An illustration of the opening angle visualization based on two functions. Both functions (distance functions based on two different centers) are indicated by contour lines. The Pareto set based in these two function is colored in green while the opening angles are visualized as a grayscale map.

Comparing this image with Figure 7.3 in Section 7.1, where ascending and descending cones are simply shown with a constant color over a white background, the comparable area around a point. Note that for the remainder of this section, if a grayscale legend at the right of images exists, it corresponds to α^+ , while the images uses the quantized version β^+ . Decomposing the domain based on β^+ meets one of the goals defined for this application, since an additional visual cue for the *robustness* of non-Pareto-optimal regions is added to the map. The α^+ -map not only shows how similar the individual fields are to each other, but also if small changes in the field gradients can transform a regular point into a Pareto-optimum, hence the term of robustness.

As a final remark note that the α^+ -map can also be defined for any d -dimensional simplicial complex, where it would, for example correspond to the intersection size between the ascending set and a d -dimensional globe.

7.3.3 Glyph-Based Visualization of Field Contribution

While the domain decomposition, i.e. the α^+ -map, enables a closer exploration of the multifield structures of a data set, the contribution of individual fields to the degree of Pareto optimality is not visualized, especially not inside Pareto-optimal regions.

To address this issue and thus the second application goal, a glyph-based visualization technique is utilized. In contrast to color maps or graphs, glyphs can condense high-dimensional information for discrete regions through single icons [15, 39]. They are regularly used in the context of tensor fields visualization [101], but are also applied for climate analysis [142] and topology [147].

As a reminder, the gradients ∇f_i inside a triangle are constant for each field. Thus, each triangle can be assigned a multi-vector equivalent to the gradients at its centroid. Furthermore, using the definition $H_\sigma^+(x)$ from Section 4.1 and results from

Section 6.2, i.e. that a point is Pareto optimal iff the gradients are linearly dependent with the scalar restricted to non-negative values, it is reasonable to provide information about the Pareto extrema and regular points based on their multi-vectors.

This is done through tailored glyph design based on work by Hüb et al. [80]. This design is used to visualize the gradient direction for each field, clusters of similar vectors, representatives of these clusters, and outliers. Outliers, i.e. single vectors with deviating directions from the clusters, indicate local dissimilar behavior of the corresponding projection and climate model. Hence, the glyphs are designed such that representatives and outliers can easily be identified while clutter from multiple similar vectors is avoided.

Glyph Design

For the visualization, alternative glyph designs are considered but rejected for the final implementation since they do not match the requirements described above.

Based on an evaluation by Fuchs et al. [61], the work focuses on radial glyphs are considered here, since they provide an effective way to representing multifield data in their spatial context. The vectors can either be shown directly in a cyclic glyph or indirectly, for example through density distribution. Both design ideas are shown in Figure 7.17 with the distribution visualized both continuously and discretized in a commonly used clock glyph [60]. Clock glyphs and pie-charts [144] subdivide a circle into several cones of equal size and, in the example, the gray scale indicates the density of vectors in each cone. This provides a good overview of the general directions in the vector set as well as clusters of vectors pointing in similar directions, and is therefore suitable to compare a set of vectors. Hence, this design can also be useful to explore the separated regions of the domain based on the similarity of the glyphs.

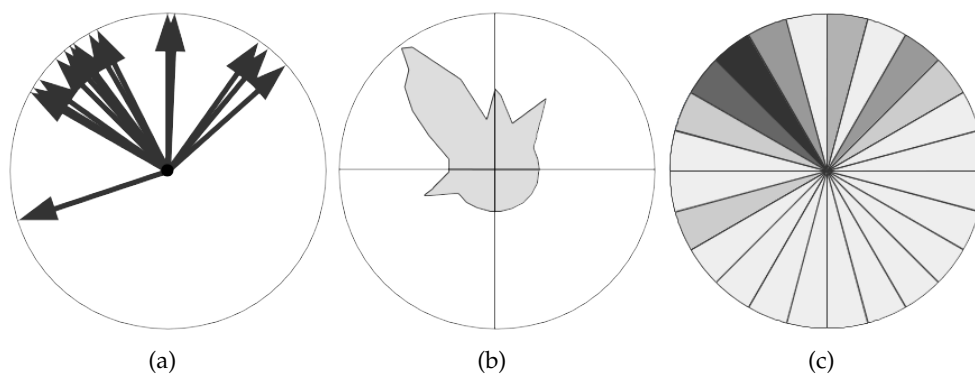


FIGURE 7.17: Three glyph designs that could be used to visualize the directions in a vector set. However, none of these designs can highlight outliers sufficiently well.

However, such indirect encoding does not provide information about specific vectors such as outliers or representatives for groups of similar vectors. Furthermore, depending on the glyph size, the existence of outliers might be obscured.

Hence, a direct encoding of the vectors can be used for an arrow-based glyph designs [4]. Figure 7.17(a) shows such a design. Note that a simple color code could be applied to identify the individual fields under investigation. For clusters of similar vectors, however, occlusion might distort the color perception and result in wrong identification.

Therefore, Häb et al. added an angle offset p to the arrow-based design. Each vector v is transformed into a minor circular sector with the angles between v and the two radii of the sector set to p and $-p$, respectively, resulting in a central angle of $2p$ for the circular sector itself.

If two sectors overlap due to similar corresponding vectors, the overlapping region is removed by reducing the size of both sectors. The two radii bordering that region are moved to the center of the overlap while the other two radii of the sectors stay the same, so that both sectors are defined by a new adapted angle offset p' towards the former overlap, and the previous offset p in the other direction [80]. See Figure 7.18 for a schematic illustration of this pairwise overlap removal.

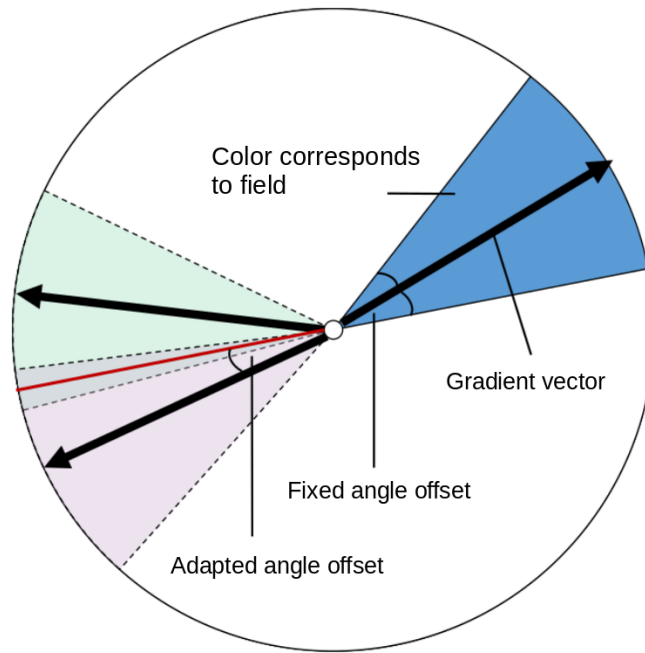


FIGURE 7.18: Modified illustration from [80] of a glyph for three gradients.

Note that for multiple vectors with multiple overlaps, any sequence of pairwise removal operations result in the same set of distinct circular sectors.

With this approach, fields with similar gradients and thus large overlaps of the corresponding circular sectors are transformed into small slices after the overlap removal. In contrast, large sectors without overlap visually highlight outliers, i.e. the value distribution of the field corresponding to that sector is different to that of the remaining fields. Thus, the size of the sector visually encodes how distinct the included fields are from each other.

For clusters of similar vectors, overlap removal yields the vectors at the cluster borders to be visualized as larger sectors than those inside the cluster. This automatically highlights possible representatives of those clusters by making the sectors defining their borders more prominent than those inside. These vectors can be called the representatives of a cluster, since for regular points, they represent the defining indices (see Subsection 4.5.2).

Averaged and Detailed Glyphs

To provide an overview of the components contributing the domain decomposition described before, a single vector set is created for each component. Each vector in this set is the averaged result of all vectors corresponding to the same field over all triangles inside the component. Hence, each set contains n vectors $v \in \mathbb{R}^d$ with n the number of individual fields and d the domain dimension. The resulting *averaged glyph* is then placed at the centroid of all involved triangle centers. The averaged glyph is depicted as an overlay over the spatially contiguous region corresponding to the connected component. To avoid visual clutter, the glyph only appears when a certain region is selected.

While the glyphs provide an overview over local topological structures, averaging over a field's angles might visually distort the impression of the actual angles. To compensate for this issue, for larger components, a further decomposition is applied based on a regular grid over the whole domain. This is due to the fact that Pareto optimal regions cannot be further separated by refining the decomposing function β^+ . This grid refinement can be done continuously or in discrete steps depending on processing power and domain size. Obviously, the finest resolution is to replace the average glyph representing an entire region by the glyphs corresponding to the triangles belonging to the selected connected component. These glyphs, based only on the multifield values inside a single triangle, are denoted as *detailed glyphs*.

Note that in this design no further information such as, for example, the scalar values of individual fields, is added. The reason for this is that information based on the scalar values might be counter-intuitive to the opening angle visualization surrounding the glyphs. It is furthermore reasonable to assume that, in most cases, the length is not distinguishable in the glyphs for the most detailed resolution (compare Figure 7.27). As for the averaged glyphs, however, it could be advantageous to relate the sector lengths to statistical measures such as the variance of the summarized vector set.

7.3.4 Scenario Overview: Global and Regional Climate Ensembles

For projections of climate change, researchers make use of a multitude of simulation runs to assess the uncertainties that are associated with the projection of key variables. Following this goal, the enhanced Pareto set visualization and the introduced glyph-design are applied to projection of average near-surface air temperature differences (ΔT_{as}) between a reference period (1961-1990) and the end of the 21st century (2070-2099). The average over the 30 year periods, a time-range over which climate is defined [154], generalizes weather information and filters out small time events and seasonal effects.

On the global level, the analyzed climate projections are conducted using the MPI-ESM model, which is part of the CMIP5 multi-model ensemble [68]. A set of four projections are considered, all forced by the same emission path RCP8.5, but differing in the horizontal resolution of the ocean component and in the number of vertical model layers of the atmospheric component, later indicated by the suffixes 'LR' (low resolution) or 'MR' (medium resolution). Figure 7.19 shows color maps for ΔT_{as} of the four individual fields. The images and similar color maps shown in this section are provided in cooperation with K. Feige through *R* code [136].

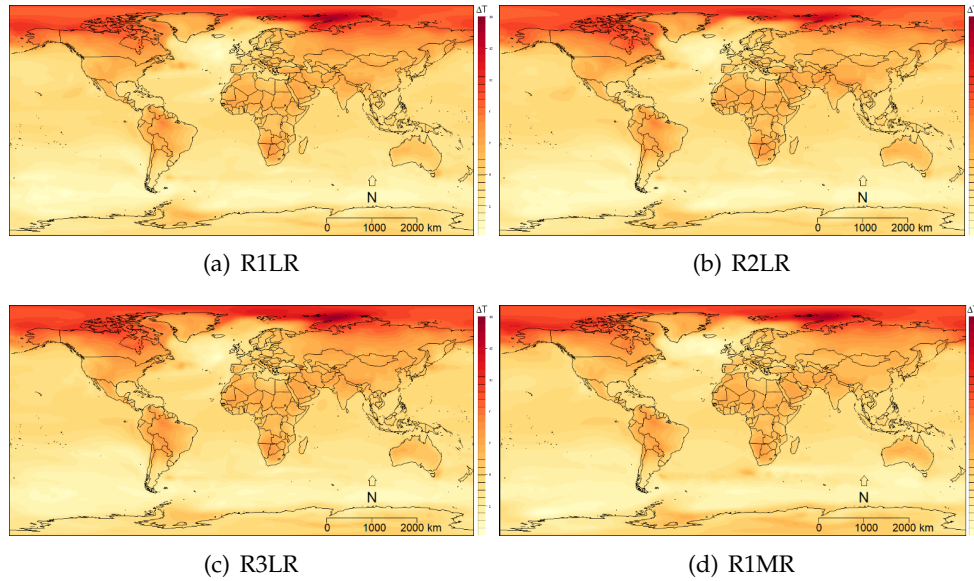


FIGURE 7.19: Each image shows ΔT_{as} for a different climate projection, as indicated by the image label. While the overlay of a world map provides spatial context, comparison between different projections is obviously difficult using only such a simple multi-window view.

On the regional level, agreement and disagreement between several subsets of the *Downscaled CMIP3 and CMIP5 Climate and Hydrology Projections* [22] are explored. The data in the archive originate from global climate ensemble simulations, varying in terms of the applied modeling system, the employed initialization strategies and other boundary conditions. These simulations, gathered by the World Climate Research Program (WCRP), specifically from phase 3 and 5 of the Coupled Model Intercomparison Project (CMIP), are then downscaled to the contiguous United States. For this study, gridded monthly average near-surface air temperatures (T_{as}) at a spatial resolution of $1/8^\circ$ from the archive [22] are used, but limited to those projections conducted by the models CSIRO-MK3-6 (Commonwealth Scientific and Industrial Research Organization, Australian), CCSM4 (National Center for Atmospheric Research, USA), and CanESM2 (Canadian Centre for Climate Modelling and Analysis, Canada) during CMIP5. Note that the Pareto sets are computed and visualized individually for each model family.

For all ensembles regarded here, ΔT_{as} is given as point data in a Cartesian grid with two dimensions. Each grid cell is split into two triangles, resulting in a connected simplicial complex \mathbb{S} without holes. The data, given at the vertices of \mathbb{S} , are

extended to multifield functions by using barycentric interpolation $f = (f_1, \dots, f_n)$. Hence, each underlying function $f_i : \mathbb{S} \mapsto \mathbb{R}$, $1 \leq i \leq n$ is piecewise linear and corresponds to one specific projection.

7.3.5 Results

Global Climate Projections

For the analysis of the global-scale ensemble, the Pareto set visualization is applied in two sequential steps. First, regions are identified where both a classic scalar-based method as well as the Pareto set approach imply a divergence between the climate projections. Then, in a second step, the glyph visualization is used to explore, which projection is responsible for these features.

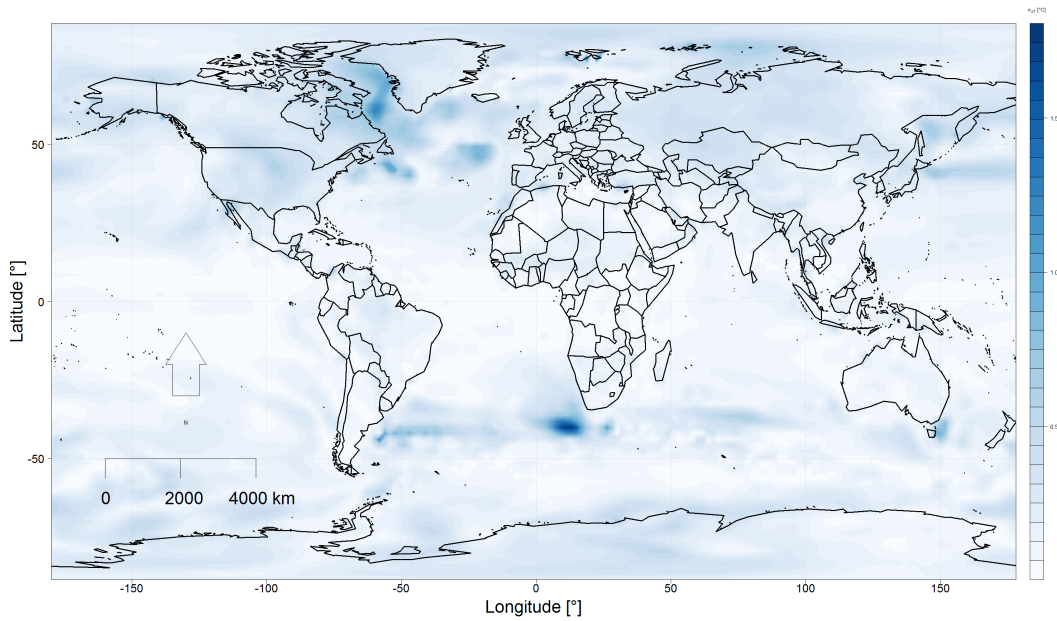
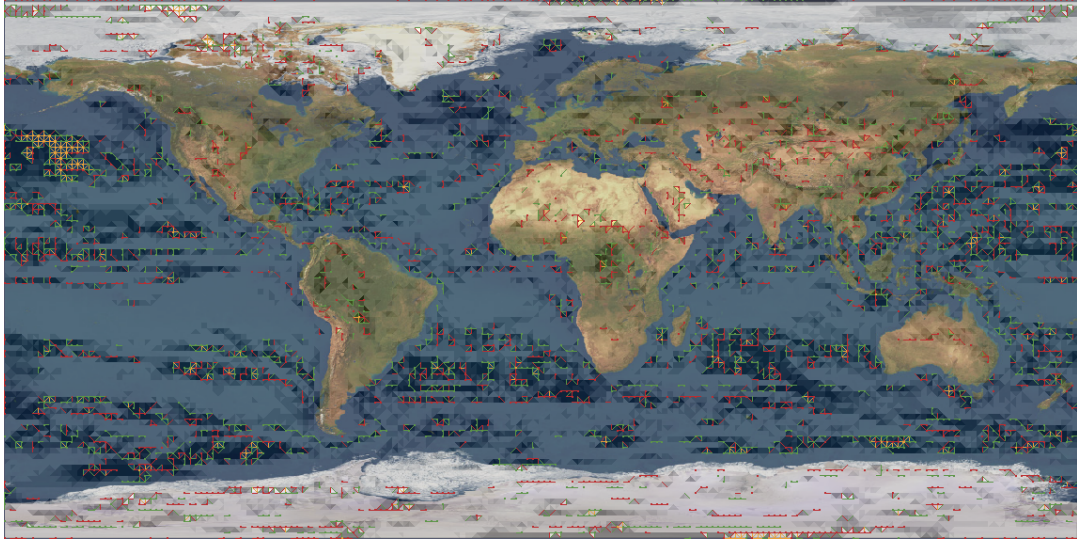


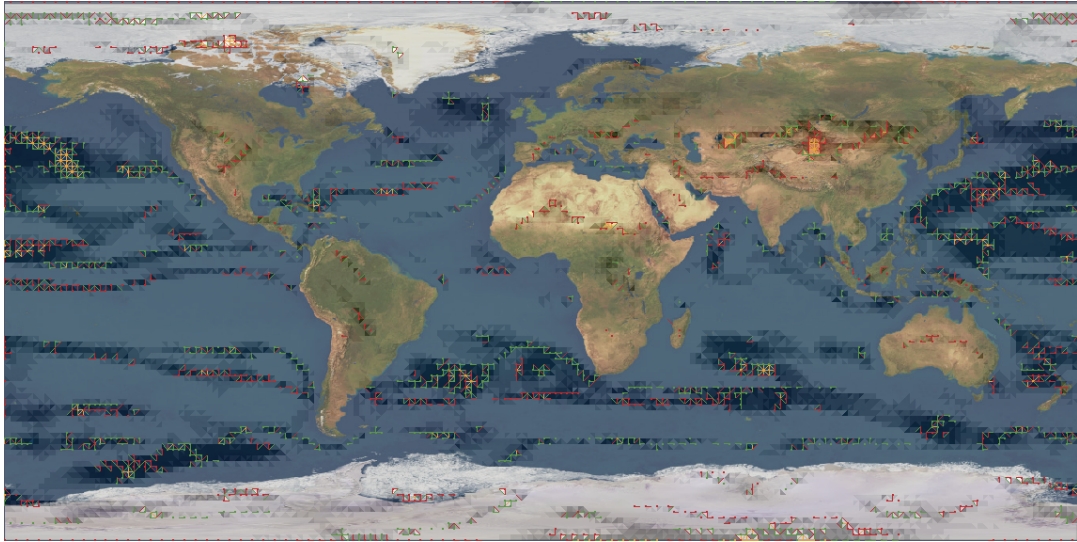
FIGURE 7.20: The ISD-map for four global climate projections based on the MPI-ESM model over a 192x96 grid.

As a scalar-based method, the standard deviation of ΔT_{as} , denoted $\Sigma \Delta T_{as}$ is presented in Figure 7.20. For the Pareto set visualization, the projections are preprocessed with one iteration of a Gaussian smoothing over a distance of three neighbors per triangle. Using this smoothing step, high spatial frequencies and purely local structures are removed to focus on the general trends within the data, which would otherwise be difficult to detect. This is due to the fact that for the Pareto set calculation, the noise from the individual projections are commutative. Since the smoothing only takes a small neighborhood around each point into account, it only removes minor components of the Pareto set as shown in Figure 7.21, where the results belonging to a smoothed and not smoothed data set are contrasted.

With both Figures at hand, it is possible to identify four types of regions, containing both Pareto sets and high $\Sigma \Delta T_{as}$ -values, either or none. As stated above, of primary interest are regions where both approaches indicate an important feature.



(a) No smoothing



(b) Smoothed

FIGURE 7.21: A world map with the Pareto set visualization as overlay. The opening angles in each triangle are presented as a transparent grayscale map, with dark gray indicating a small opening angles, and thus high divergence between the projections. Note for example, that incomparability mostly appears over the oceans.

To better identify those regions, both images are merged into one Figure 7.22 using additional visualization tools (ParaView and R). In this figure, the background contains the map of $\Sigma\Delta T_{as}$ from Figure 7.20. As compared to Figure 7.20, the color saturation is increased to compensate for an overlay, which shows Pareto extremal areas colored red, green, and yellow depending on their Pareto status and the opening angle as a transparent grayscale map. Note that with this overlay, a 2D color space is created with white to blue in one dimension and white to gray in the second dimension.

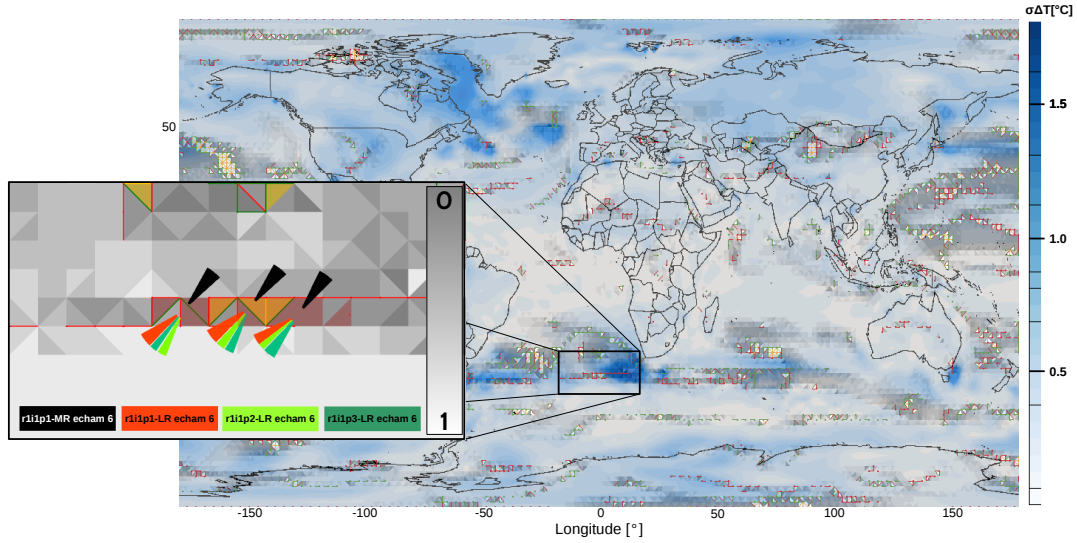


FIGURE 7.22: A merge between Figure 7.20 and the smoothed Image 7.21(b). The background contains the map of $\Sigma\Delta T_{as}$ from Figure 7.20, while the Pareto-set visualization is presented as an overlay.

With this visualization, one can, for example, focus at a region near the south end of Africa, based on the dark blue color in the $\Sigma\Delta T_{as}$ layer, and nearby Pareto-extremal regions and high opening angle values (indicated by darker transparent gray areas near the Pareto extrema) that appear in the overlay. Other regions, where only one of the two approaches indicate a large feature, for example in the Philippine Sea for Pareto sets or between Greenland and New Foundland for the ISD-map, might also be interesting, but the analysis is focused to features present in both approaches. Note that in the case of the Philippine region, one can assume that appearing Pareto extrema correspond to the vortices created by the Kuroshio current (see Figure 7.23), part of the North Pacific ocean gyre [93, 85, 7].

In the second analysis step, the glyphs are used for the focused area by selecting components of the Pareto set or sections resulting from the domain decomposition. The area detail shown in Figure 7.22 visualize the Pareto sets colored red, green and yellow and, as a grayscale map, the opening angles. Exemplarily, three sections near the high $\Sigma\Delta T_{as}$ values seen in the overview images are selected, which is highlighted by slightly darker colors and associated glyphs in the Pareto set visualization. The colors of each glyph segment correspond to one of the four climate projections:

- r1i1p1-MR_echam6** in black,
- r1i1p1-LR_echam6** in orange,
- r2i1p1-LR_echam6** in green, and
- r3i1p1-LR_echam6** in turquoise.

Note that the black segment is isolated in each of the three glyphs, indicating an outlier. The black segment corresponds to the MPI-ESM-MR projection, which differs to the LR-projection amongst others in a higher resolution over the ocean tiles,

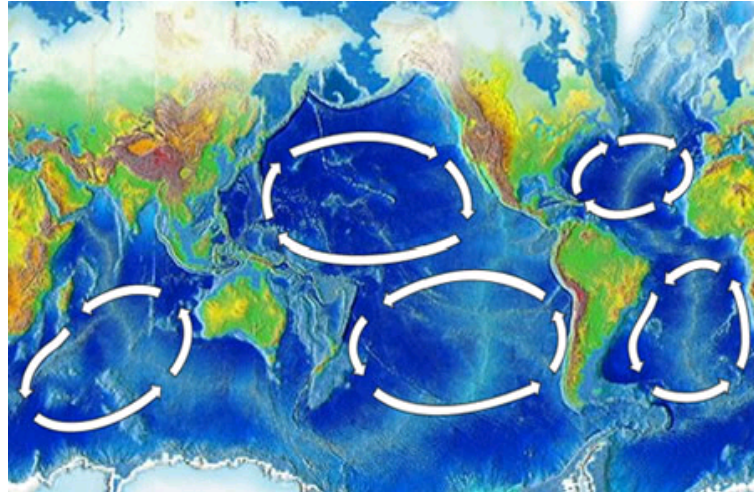


FIGURE 7.23: The Kuroshio current is the west side of the clockwise North Pacific ocean gyre. The image is from the free media repository from the Wikipedia website. [183]

which makes it eddy-permitting. This is a structural model difference, which results in a different capability to resolve ocean eddies which could be responsible for the diverging climate signal projected here.

Regional Climate Projections

In the second application example, the uncertainty within a set of downscaled climate projections over the contiguous United States is analyzed. A preliminary calculation of the Pareto set for all 70 projections included in the entire CMIP5 ensemble showed that the topological differences between the ensemble members are too strong to analyze them as a whole. On the other hand, the differences between the gradients in the predicted average near-surface air temperatures, regarded separately for both time windows (reference period (1961-1990) and end of simulation period (2070-2099)), are each very small. Hence, three smaller, but still reasonable subsets of the entire ensemble are chosen, whereas each subset corresponds to simulations conducted using the same model (CSIRO-MK3-6, CCSM4, and CanESM2) and emission path RCP8.5. Thus, only the uncertainties due to different initial conditions are visualized and investigated. The separate set contain 5, 5, and 10 projections, respectively.

Similar, but not completely consistent behavior can be found in all three subsets, such as large, topologically regular regions in the south east of the USA and clusters of Pareto extrema in and around the Rocky Mountains. This can be observed in the examples (d) - (f) shown in Figure 7.24 and compared to the standard deviation map for the same subsets in Figure 7.24 (a) - (c).

Again, the $\Sigma\Delta T_{as}$ -values are overlaid with the Pareto sets and opening angle visualization, as done in the previous section for the global-scale ensemble. This enables us to identify similarities and differences between these approaches, exemplarily shown for CSIRO-MK3-6 in Figure 7.25(a). Note, for example, the similarity between both visualizations near the Canadian border around North Dakota. However, since both techniques are different – the standard deviation is computed for isolated points, while the Pareto set concept is based on gradient directions – one can

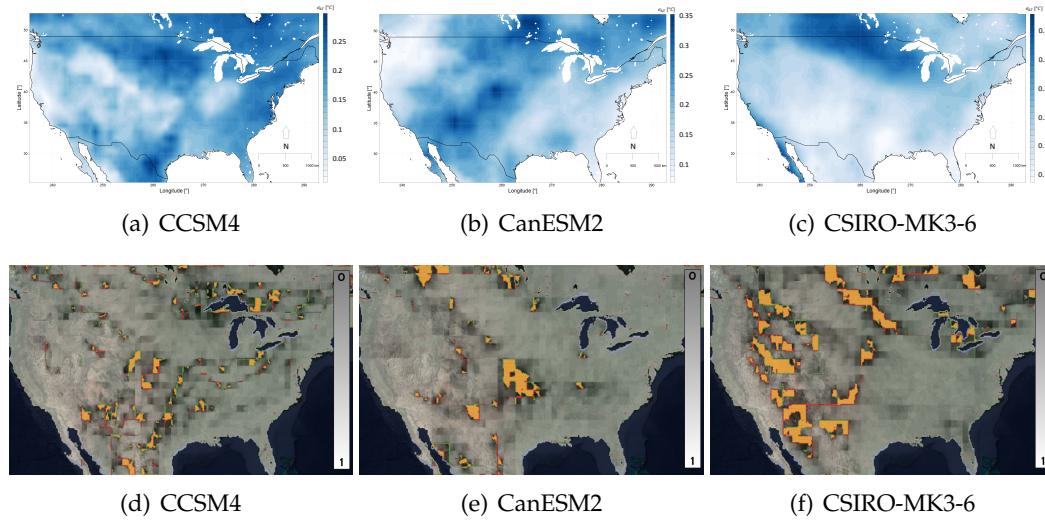


FIGURE 7.24: The first row, images (a)-(c), show the ISD-map for different subsets of projections based on the same climate model indicated by their subtitles but with different initial conditions. Images (d)-(f) show a world map and the Pareto set visualization for the same subsets.

also find large differences in their identified features, for example along the mountain regions in the eastern USA.

In contrast to the global-scale results discussed before, the connected components of Pareto extrema are larger and more numerous for the regional-scale ensembles. This leads us to the assumption that the projections that are responsible for these dissimilarities might change throughout components. To analyze the effect of single projections towards the Pareto set visualization, an alternative approach than the one used for the global ensemble is used. Therefore, the data set is visualized with the glyph-based approach and a β^+ -map instead of the ISD-map. A possible option is to select several regions with high topological divergence, i.e. components of Pareto-optimal regions or simply regular regions showing low opening angles, as demonstrated in Figure 7.25(b) and 7.26. In each image, the glyphs for the selected regions are calculated and, based on their visualization, one or two projections are removed to analyze their contribution to topological divergence.

According to the two glyphs in Figure 7.25(b), in both selected components the projections CSIRO-MK3-6-0. i with i from 8 to 10 might be responsible for the large size of the Pareto set since each of the three projections seem to be an outlier in at least one of the glyphs. Figure 7.26 shows the section of the same domain after CSIRO-MK3-6-0.9 and CSIRO-MK3-6-0.10 are sequentially removed. Note how both previously selected regions are now separated into much smaller Pareto-optimal regions.

As shown in Figure 7.26, this approach can be repeated considering even more individual regions. To avoid confusion, it is possible to highlight a specific glyph segment corresponding to the same projection throughout all glyphs in Figure 7.26(a). This facilitates the identification of regions, in which a specific projection is isolated, allowing for an overview over the contributors to a large group of Pareto extremal

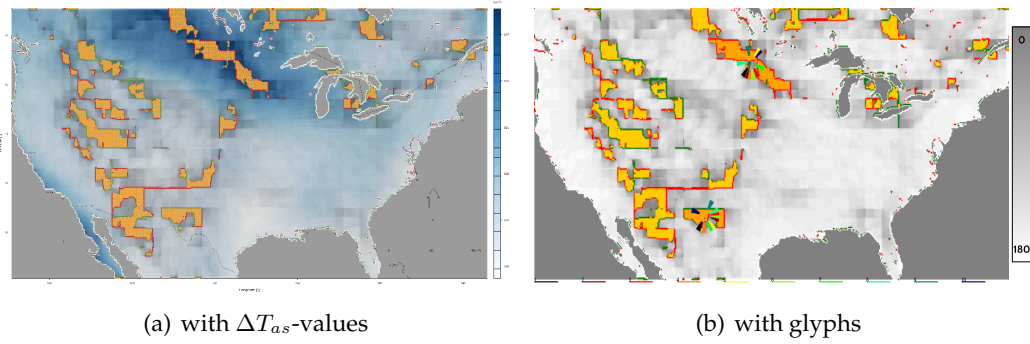


FIGURE 7.25: Image (a) shows a merge between the $\Sigma\Delta T_{as}$ -values (Figure 7.24(c)) and the Pareto sets and opening angle visualization (Figure 7.24(f)) for the climate projections based on CSIRO-MK3-6 over the United States. In (b), two larger Pareto extremal regions are selected and highlighted in a slightly darker orange, to identify possible projections, responsible for this large topological divergence. Note that the $\Sigma\Delta T_{as}$ -color map is removed in (b) to make it easier to connect the glyph segments to their corresponding projection.

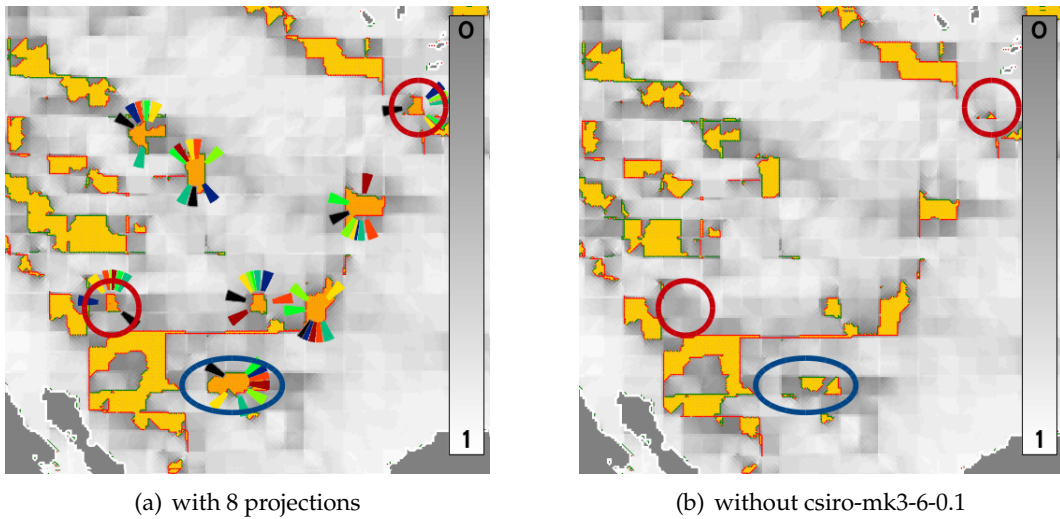


FIGURE 7.26: A section of the United States with the Pareto set and opening angle visualization for the CSIRO-MK3-6.i projections, $1 \leq i \leq 8$. In (a), a selection of eight Pareto extremal regions is chosen to analyze, which projections are responsible for the topological divergence around this section of the USA. Image (b) shows the same section after CSIRO-MK3-6-0.1 is removed from the set of projections.

Note the decreasing extent of the Pareto-optimal regions.

regions. Hence, it becomes possible to see if the disagreement between the models over a larger area can be attributed to a certain ensemble member.

The visualization of the detailed glyphs at a higher resolution, on the other hand, provides a better understanding of individual model behavior inside those regions.

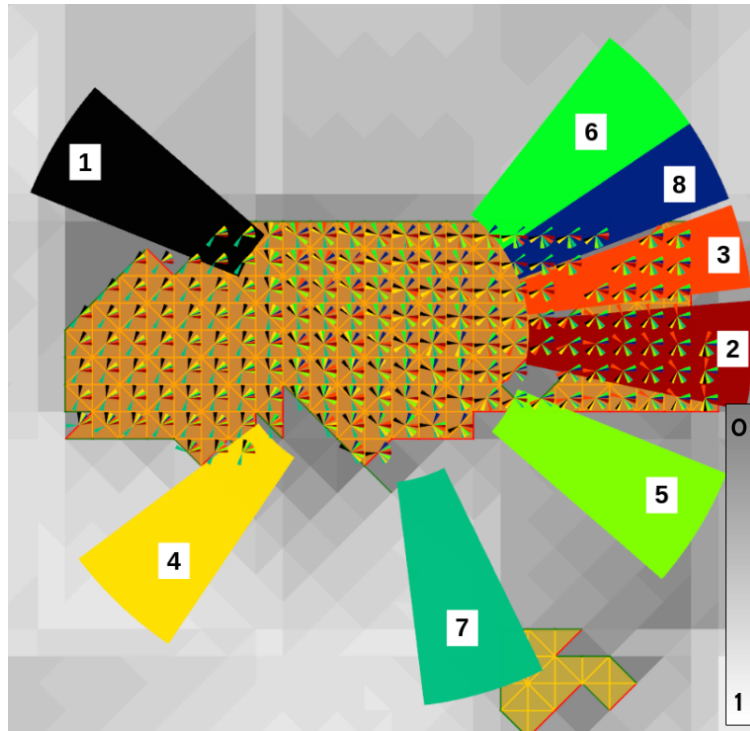


FIGURE 7.27: A detailed visualization of a region, selected in Figure 7.26(a). The numbers i in each glyph segment indicate the corresponding projection CSIRO-MK3-6-0. i .

Figure 7.27 shows such detailed visualization for one of the selected regions in Figure 7.26(a). Note how the averaging process, resulting in the overview glyph, distorted the information that the black segments, corresponding to projection CSIRO-MK3-6-0.1, is only a distinct outlier in a portion of the Pareto region. In other parts, the orange or the turquoise segments, i.e. projection CSIRO-MK3-6-0.3 and CSIRO-MK3-6-0.7, respectively, are even more distinct from the rest of the projections. This is reflected in the Pareto set shown in Figure 7.26(b), for which CSIRO-MK3-6-0.1 is removed from the set of projections. Note how the Pareto optimal regions that used to be in the red encircled areas disappeared, while the Pareto set in the blue encircled area is only reduced. While not yet implemented for this data set, further sequential insertion and removal of projections is a possible tool to study the contribution of the individual projections towards the Pareto set. This allows to temporarily insert an additional projection to the multifield and to study, for example, the change of robustness. Using this technique, regions where the added field shows a different topological behavior than the original multifield, can be identified more easily.

Larger Numbers of Projections

To present the capability of Pareto sets as a visualization technique for more than 5 to 10 projections, this paragraph considers all available projections for T_{as} for April 2020. Each field corresponds to a different modeling system employed in the context of CMIP3 based on the emission path SRES B1. For testing purposes, also a CCSM3 projection (run 4) based on the SRES A2 is included, resulting in a total of 40 different projections. The goal is to identify locations where the climate models did not produce topologically similar results. Since, the data set is only based on one

month instead of a long time period, the visualization does not provide information regarding the climate, but only regarding the different projections.

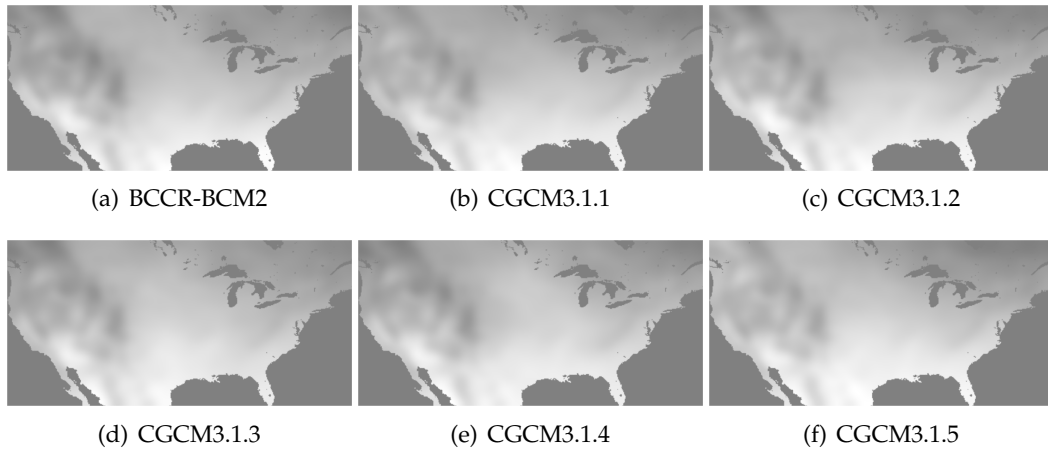


FIGURE 7.28: Grayscale map of T_{as} for April 2020 as projected by six modeling systems employed in CMIP3. References to the models can be found in [22]. Darker greys refer to lower and lighter greys to higher temperatures.

Figure 7.28 shows the distribution of T_{as} for six of the 40 climate projections as grayscale maps. The scalar fields seem very similar, making them difficult to distinguish using a multi-view display.

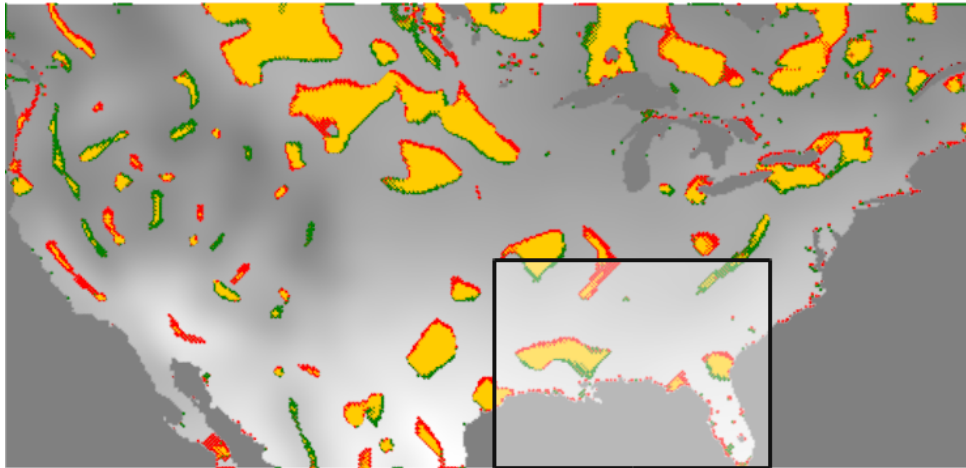


FIGURE 7.29: The Pareto set for 40 projections. The box indicates a region where further investigation is done, see Figure 7.30.

To compare the projections, the Pareto set is computed as seen in Figure 7.29. Note the dark gray coloring in the marked area covering Florida, that indicate a low robustness and, therefore, a less similar behavior among the individual fields. To further explore the contribution of individual fields to the Pareto optimal regions in the marked area, the glyph-based visualization is applied. As the results shown in Figure 7.30 b-f suggest, a small subset of projections disagree with the majority of the

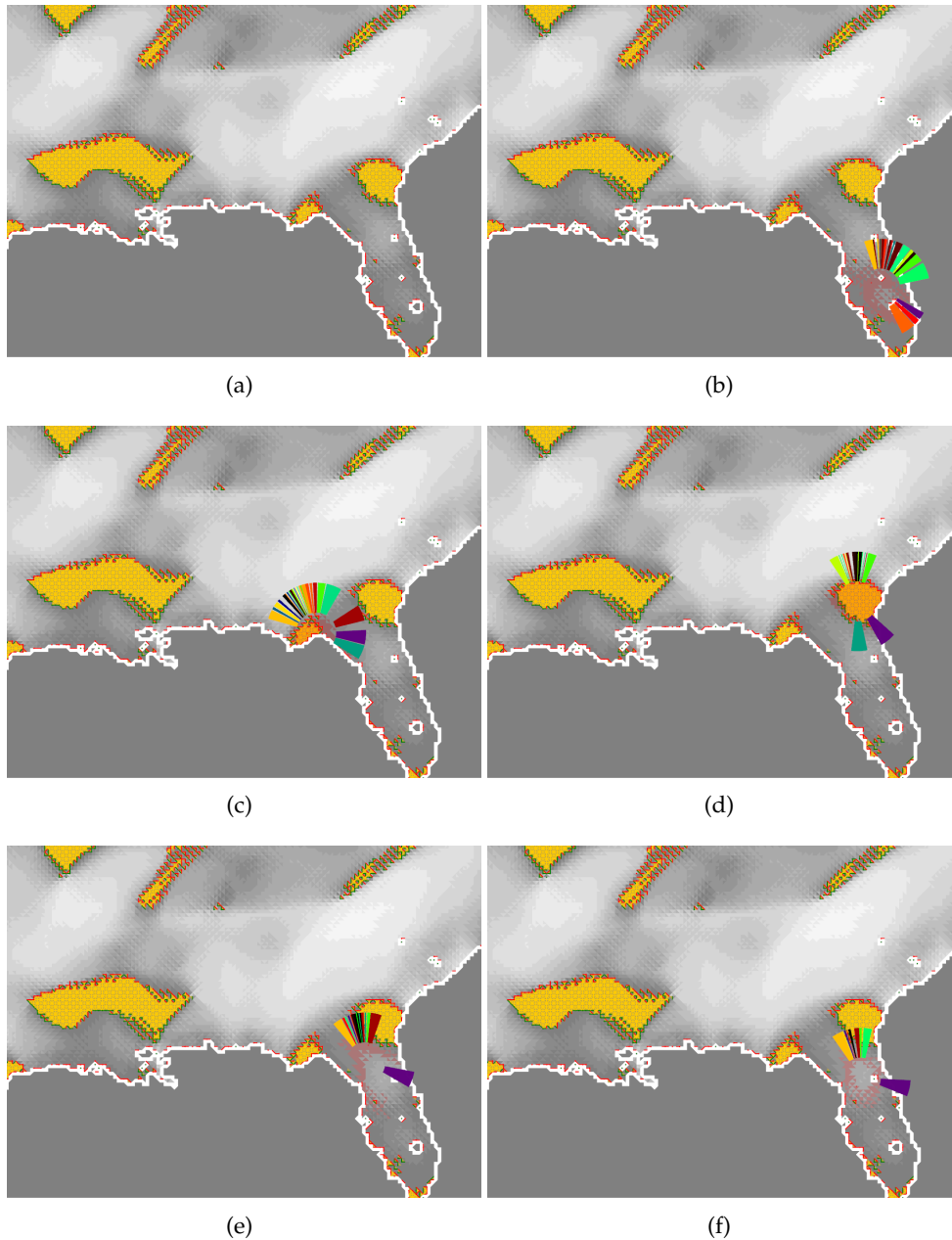


FIGURE 7.30: Using glyphs to explore diverging model behavior.

simulations. In particular, the violet (HadCM3, run 1) glyph sector exhibit distinct orientations in all shown examples, while the turquoise sector (CCSM3, SRES A2, run 4) also stands out in two of the shown cases. As already assumed above, the CCSM3 projection corresponds to the artificially included outlier, which is based on a different emission scenario than the remaining projections.

Note that the computation of the Pareto set and the α^+ -map is local, i.e. is done for each simplex separately. Hence, with the implementations outlined in Section 4.5, the calculations should, in theory, scale linearly with respect to increase numbers of fields and simplices. To support this assumption, that the runtime to visualize the Pareto set remains reasonable even for large numbers of fields, Table 7.2 lists the

Year	Data set		Incremental [sec.]	Direct [sec.]	MT [sec.]
	#Var	CMIP			
2016	20	3	592	311	2625
2016	40	3	1176	551	3237
2011	39	3	1256	484	3194
2011	70	5	2312	924	7913
2020	39	3	1273	502	3238
2020	70	5	2255	848	7807

TABLE 7.2: Data set characteristics and run time in seconds for different Pareto set calculations

runtimes for different Pareto set algorithms, applied on the data sets similar to those presented in this section, in detail the averaged air temperature over the United States. Each data set contained the averaged value over the month April in the year indicated by the first column. The second column (#Var) indicates the number of fields included in the different multiple data sets. All fields are projections based on models using different emission others and either CMIP3 or CMIP5, indicated by the third column. The last three columns present the runtime in seconds. The first algorithms calculate the Pareto set sequentially, i.e. the Pareto set is recalculated after each field is added sequentially, saving and using the defining indices. The second one, denoted *Direct*, calculates the Pareto set only after all fields are loaded, still using the defining indices but without the computational overhead of initializing the sequential calculation process. Thus, this runtime does not include repeated space allocation or function initialization, but it also excludes the capability of visualizing interim results. The last column of the table shows the run time using the marching triangle method, which is here denoted as MT. The implementation guidelines for the algorithms can be found in Section 4.5. For the computation, an Intel Core 2 Quad CPU with 2.40GHz per core and 8GB memory is used.

The table shows that a main obstacle in the current implementation is the amount of computational overhead. Also note that the time in all columns more or less doubles if the number of individual fields doubles; slightly more for the MT-column and slightly less for the other two.

Discussion

In this application scenario, several improvements towards the analysis and visualization of multifield data through Pareto sets are added.

- The domain is decomposed into regions based on similar field behavior / robustness. The used α^+ -map can be combined or compared to common visualization methods such as the ISD-map.
- To visualize/analyze individual fields and their contribution to the Pareto set, the applied domain decomposition is complemented by a glyph-based visualization. With this approach the identification of fields behaving similarly is eased, but also highlights corresponding outliers.

The results of this application imply several tasks for future work. First, the integration of variance information into the averaged glyphs might be advantageous in cases, where the averaging distorts the actual gradient directions. On the other

hand, if the distortion is only due to averaging over larger region, this can be solved by either refining the decomposition or by additionally separating the regions, for example, based on a regular grid.

Second, using the glyphs, regions might be further separated and visualized, on top of the separation method based on opening angles. Distance measures between the used circular glyphs are already at hand, for example in work by Hüb et al. [79], or alternatively, the gradient set approach by Nagaraj et al. [122]. Both can be used with standard clustering algorithms to separate the domain and to provide an overview over larger sections of the domain. This can even be extended to compare different combinations of multi-fields, such as, e.g., projections that are grouped by certain criteria.

Subsection 7.3.5 shows that the approach also works for large numbers of individual fields. However, in the context of climate studies, i.e. data over long-term periods like $\Sigma\Delta T_{as}$, the Pareto set approach produces better visual results for multifields containing at most 10 fields and are created with the same climate model. A possible explanation for this limitation is that, while climate models produce topologically similar short-term projections, the created long-term climate projections are rather unrelated when compared to projections from other models. This assumption needs to be further investigated and shows another example, how the Pareto set visualization opens and guides future research topics.

In summary, the improvements in this section provide an exploratory tool that allows users to closely investigate the composition of multifield data sets, identifying topologically interesting regions, and showing why those regions are interesting. The additional information about the individual fields displayed by the glyph summaries is also a first step to solve some of the open tasks, laid out in the previous Section 7.1 and 7.2, like the visualization of the contribution of individual fields to the Pareto set.

Chapter 8

Conclusion

In this thesis, concepts from multiobjective optimization are adapted to define extremal points in multifields and thus to identify regions of similar or dissimilar topological behavior among the individual fields with respect to ascending and descending directions. All necessary definitions and algorithms are presented to implement, extend or improve the Pareto set concept and also its interior structure, visualized through the reachability graph. Common multifield obstacles like *noise* and data complexity are discussed and handled through simplification approaches.

Examples and application scenarios are provided and processed to show the interpretation, utility, and potential of the Pareto set in visualization applications. While describing those scenarios, further improvements of the Pareto set are presented as well as its integration with other visualization tools. Improvements and extensions include:

- the weighted Pareto set, that assigns weights to the connected components inside the Pareto set based on extremal points in the individual fields and thus allow to assess the relevance of the different components in the considered application scenario,
- the *robustness* term, that provides a measure of how close a regular point is to become Pareto extremal due to small changes in the values of the individual fields and thus provides a continuous decomposition of non-Pareto optimal points, and
- the glyph-based visualization of regular and Pareto optimal regions based on averaged gradients to visualize outliers and subgroups of similar fields in those regions with respect to their gradients.

The Pareto set, in combination with the reachability graph and the listed improvements allow for a simple visualization and exploration of scalar multifield data sets to support domain experts in a variety of applications.

As there are currently several concurrent approaches tailored to the topological study of multifields, a detailed discussion of similarities and differences of the Pareto approach to these methods is provided. The corresponding sections prove the close subset relationship between the Pareto set and the Jacobi set, assuming a restriction of number of individual fields by the domain dimension, and present a parameterized approximation of the Pareto set through the directed Joint Contour Tree that directly implies a similar relation with the Reeb space topology. This is supported through related work by Wan [169], in which the author define Morse theory

for two fields based on Pareto optimality. However, the concept as presented in this thesis remains topology-*based* which is sufficient for visualization and application purposes.

8.1 Discussion, Limitation and Relevance

With the Pareto set concepts presented in this thesis, it is possible to handle continuous and, as this is particularly relevant for practical applications, piecewise linear multifields over simplicial complexes with an arbitrary number of individual fields without restrictions to the domain dimension. This is especially vital if the scenario contains far more than 10 or 20 individual fields. However, based on the definition of Pareto optima, once a point is Pareto optimal, i.e. surrounded by incomparable neighborhoods, considering additional fields cannot reverse it to a regular one. Hence, increasing the number of fields will never decrease the size of the Pareto set. Therefore, it is required that the individual fields are related and to a small degree similar, which is, for example, the case for ensemble data sets that are based on repeated simulation runs. If the fields included into the analysis become too different, it is possible that the entire domain becomes Pareto optimal. While a naive visualization approach would lack any information about which fields are responsible for Pareto-optimality, adding corresponding glyphs provides these details (see Subsection 7.3.3).

In general, the Pareto set can be directly visualized onto the geometry of the domain; a feature not available for concepts like the Reeb space. This allows for a domain decompositions, for example based on robustness (compare Section 7.3.2). Also the combination with other visualization methods, geographical information and glyphs is possible.

In contrast to statistical measures such as average, standard deviation, especially ISD, or correlation, the fields are only compared by their topology and gradients such that the Pareto set is invariant to scaling factors or a variable's unit. However, the results remain sensitive to field inversion, i.e. the Pareto set changes if instead of a field f_i , its inversion $-f_i$ is considered. This detail makes Pareto sets very suitable to visualize maximal and minimal behavior among multiple different scalar fields and also distinguishes between these two types of extremal behavior. This is a feature that previous definitions of Pareto optimality, for example in the work by Smale [150], do not include.

Additionally and in contrast to existing approaches like JCN, Jacobi sets, etc., directional information of the individual fields are incorporated into the analysis. This not only allows for the introduction of ascending and descending sets in the multifield representation but it also makes the definition of robustness and the utilization of the presented glyph design possible.

The directional information is furthermore used to define the reachability graphs, an approach to visualize the global structure of the Pareto set along with the connections between its subregions. This inspired the definition of the dJCN (Definition 6.3.1 in Subsection 6.3.3). Note that like JCNs, contour trees, and Reeb graphs, the nodes of the $RG(f)$ are not directly linked to geometrical positions in the domain. Like the contour trees [174], different multifields f, g can have the same reachability

graph $RG(f) = RG(g)$ (appropriately node renaming assumed). Hence, following work by Harvey and Wang [82], this can be used to visualize high-dimensional multifields.

Besides naive approaches like changing the rounding function for JCNs, this work is amongst one of the first to address the simplification of scalar multifield structures. Although other simplification approaches, for example the Jacobi sets, already exist, they pose restrictions on the number of fields and dimensions and are invariant to gradients. The ideas introduced in this thesis work in theory and smaller examples, but are not yet scalable to larger data sets, as they are limited by the high computational effort to create the reachability graph as a support structure.

However, excluding the reachability graph and focusing only on the Pareto set, the implementation is easily parallelizable, since Pareto extremity is calculated locally. Therefore, the current implementation already uses multiple threads to compute the Pareto set. With respect to increasing field numbers, the concepts and the computational effort of their implementation scales well for 2D domains. For higher-dimensional domains, however, scaling is limited because ascending and descending sets are calculated through the marching tetrahedrons algorithm for the three-dimensional case, or linear programming for domain dimensions higher than three. For two dimensional cases, the approach using defining indices (see Subsection 4.5.2) scales linearly with increasing field numbers. The same scalability holds for the domain size, i.e. the number of simplices. As a reminder, the implementation is only outlined for two- and three dimensional domains due to the application-orientated focus of the thesis.

8.2 Outlook

While the mentioned limitations are open tasks itself, additional open questions, and possible improvements and directions for future work are mentioned throughout the dissertation. Those include:

- While this work is based on topological notions and is ultimately aimed at describing a concept of topology for multifields, the visualization so far can only be classified as topology-based. A mathematically detailed and precise definition of multifield topology and a detailed investigation into its properties is therefore a high priority for future work.
- As mentioned in the previous sections, simplification is necessary for multifield data. So far, the presented approach is sound but rather inefficient. Additionally, in Chapter 4, small irregularities are encountered at the Pareto set borders: Here, sometimes small sections occur as Pareto maxima, although they should be Pareto minima according to the included functions, and vice versa. These errors are attributed to the fact that continuous fields are triangulated onto simplicial complexes. Both efficient global simplification and efficient local removal of those irregularities are an open task.
- Since Pareto optimality can be formulated for the continuous case, a limitation to simplicial complexes – which is considered throughout all chapters of this thesis to allow for a concise presentation and implementation – might be unnecessary. This would make it possible to analyze a larger variety of multifield

representations using the Pareto set concept. Furthermore, with the alternative definitions of Pareto extrema based on ascending and descending paths, it might also be possible to apply the concept to vector fields, similar to the idea of Morse decomposition.

- Chapter 7 shows several scenarios, in which Pareto sets could be a useful analysis tool. It is also shown, how this technique could be combined with additional approaches. However, topology-based approaches for multifields are not yet used as regular as their counterparts for single-field data sets. Hence, more work is necessary to improve the presence of Pareto sets and other concept in the research community around multifield analysis and visualization.

In summary, in this dissertation, a full and complete description of the Pareto set concept is presented as ground work for future applications of multifield data analysis and is open for a wide variety of future work and improvements. These improvements can extend the basic approach or can be tailored to individual application scenarios. Possible scenarios include those presented in the application section, but may also be found in a wide range of research and industrial areas relying on uncertainty, time-varying, and ensemble data sets in general.

Hence, to answer the research question of this dissertation, the Pareto set concept is capable to visualize multifield data in different applications and has close relationships towards other topology-based approaches, but also its own specific advantages and drawbacks. Next to visualization, the concept can be extended to support the detailed exploration of multifield data sets. In this thesis, definitions and efficient implementation guidelines are provided to create the visualization and support the exploration, and enables the research community to easily enhance the Pareto set concept. Possible improvements range from theoretical to practical levels, making the concept even more valuable such that the Pareto set concept, after this thorough introduction, can be considered among one of *the* topology-based tools for multifield visualization and analysis.

Bibliography

- [1] S. D. Aberson et al. "Hurricane Isabel (2003): New Insights Into the Physics of Intense Storms. Part II: Extreme Localized Wind". In: *Bulletin of the American Meteorological Society* 87.10 (2006), pp. 1349–1354. DOI: [10.1175/BAMS-87-10-1349](https://doi.org/10.1175/BAMS-87-10-1349).
- [2] H. Akiba and K.-L. Ma. "A Tri-Space Visualization Interface for Analyzing Time-Varying Multivariate Volume Data". In: *Eurographics Conference on Visualization (Eurovis)*. 2007, pp. 115–122. ISBN: 978-3-905673-45-6. DOI: [10.2312/VisSym/EuroVis07/115-122](https://doi.org/10.2312/VisSym/EuroVis07/115-122).
- [3] B. Alpern and L. Carter. "The Hyperbox". In: *IEEE Conference on Visualization. VIS '91*. San Diego, California: IEEE Computer Society Press, 1991, pp. 133–139. ISBN: 0-8186-2245-8.
- [4] G. L. Andrienko and N. V. Andrienko. "Spatio-temporal aggregation for visual analysis of movements". In: *Visual Analytics Science and Technology*. 2008, pp. 51–58. DOI: [10.1109/VAST.2008.4677356](https://doi.org/10.1109/VAST.2008.4677356).
- [5] L. Arge and M. Revsbæk. "I/O-Efficient Contour Tree Simplification". In: *Algorithms and Computation*. Ed. by Y. Dong, D.-Z. Du, and O. Ibarra. Vol. 5878. Lecture Notes in Computer Science. Springer Berlin, 2009, pp. 1155–1165.
- [6] C. Bajaj, A. Gillette, and S. Goswami. "Topology Based Selection and Curation of Level Sets". In: *Topology-Based Methods in Visualization II*. Ed. by H.-C. Hege, K. Polthier, and G. Scheuermann. Mathematics and Visualization. Springer Berlin Heidelberg, 2009, pp. 45–58. ISBN: 978-3-540-88605-1. DOI: [10.1007/978-3-540-88606-8_4](https://doi.org/10.1007/978-3-540-88606-8_4).
- [7] R. Barkley. "The Kuroshio current". In: *Science Journal* 6 (1970), pp. 54–60.
- [8] B. Baruque et al. "Ensemble Methods for Boosting Visualization Models". In: *Bio-Inspired Systems: Computational and Ambient Intelligence: 10th International Work-Conference on Artificial Neural Networks*. Ed. by J. Cabestany et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 165–173. ISBN: 978-3-642-02478-8. DOI: [10.1007/978-3-642-02478-8_21](https://doi.org/10.1007/978-3-642-02478-8_21).
- [9] G. K. Batchelor. *An introduction to fluid dynamics*. Cambridge University Press, 1967.
- [10] R. Becker, W. Cleveland, and M.-J. Shyu. "The Visual Design and Control of Trellis Display". In: 5 (1996), pp. 123–155.
- [11] S. Bergman. *The Kernel Function and Conformal Mapping*. Vol. 5. Mathematical Surveys and Monographs. American Mathematical Society, 1950. ISBN: 9780821873861.
- [12] *Bing Aerial Map*. <https://www.bing.com/maps/aerial>. [Online; accessed: 11-September-2017].

- [13] J. Blaas, C. Botha, and F. Post. "Extensions of Parallel Coordinates for Interactive Exploration of Large Multi-Timepoint Data Sets". In: *IEEE Transactions on Visualization and Computer Graphics* 14.6 (2008), pp. 1436–1451. ISSN: 1077-2626. DOI: [10.1109/TVCG.2008.131](https://doi.org/10.1109/TVCG.2008.131).
- [14] J. M. Bland and D. G. Altman. "Statistics notes: Measurement error". In: *BMJ* 312.7047 (1996), p. 1654. DOI: [10.1136/bmj.312.7047.1654](https://doi.org/10.1136/bmj.312.7047.1654).
- [15] R. Borgo et al. "Glyph-based Visualization: Foundations, Design Guidelines, Techniques and Applications". In: *Eurographics State of the Art Reports*. EG STARs (2013), pp. 39–63. DOI: [10.2312/conf/EG2013/stars/039-063](https://doi.org/10.2312/conf/EG2013/stars/039-063).
- [16] J. Bormann. "Joint Contour Nets revisited". MA thesis. Kaiserslautern, Germany: University of Kaiserslautern, 2016.
- [17] J. Bormann, L. Huettenberger, and C. Garth. "The approximation of Pareto Sets using directed Joint Contour Nets". in review. 2017.
- [18] R. Botchen et al. "Flow feature visualization using logical operators on multivariate fields". In: *International Symposium on Flow Visualization*. 2008.
- [19] M. Böttinger et al. "Visualization of 2D Uncertainty in Decadal Climate Predictions". In: *Workshop on Visualization in Environmental Sciences (EnvirVis)*. Ed. by A. Middel, K. Rink, and G. H. Weber. The Eurographics Association, 2015. DOI: [10.2312/envirvis.20151083](https://doi.org/10.2312/envirvis.20151083).
- [20] S. Boyd and L. Vandenberghe. *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004. ISBN: 0521833787.
- [21] M. Brandner and T. Thurner. "Uncertainty in optical measurement applications: a case study". In: *IEEE Transactions on Instrumentation and Measurement* 55.3 (2006), pp. 713–720. ISSN: 0018-9456. DOI: [10.1109/TIM.2006.873802](https://doi.org/10.1109/TIM.2006.873802).
- [22] L. Brekke et al. "Downscaled CMIP3 and CMIP5 climate and hydrology projections: Release of downscaled CMIP5 climate projections, comparison with preceding information, and summary of user needs". In: *US Army Corps of Engineers, US Geological Survey* (2013).
- [23] P.-T. Bremer et al. "Topological feature extraction and tracking". In: *Journal of Physics: Conference Series* 78 (1 2007).
- [24] P.-T. Bremer et al. "Analyzing and Tracking Burning Structures in Lean Premixed Hydrogen Flames". In: *IEEE Transactions on Visualization and Computer Graphics* 16.2 (2010), pp. 248–260. DOI: [10.1109/TVCG.2009.69](https://doi.org/10.1109/TVCG.2009.69).
- [25] R. Burger et al. "Interactive cross-detector analysis of vortical flow data". In: *Conference on Coordinated and Multiple Views in Exploratory Visualization (CMV 2007)*. 2007, pp. 98–110. DOI: [10.1109/CMV.2007.12](https://doi.org/10.1109/CMV.2007.12).
- [26] R. Bürger and H. Hauser. "Visualization of Multi-variate Scientific Data". In: *Eurographics State of the Art Reports*. 2007, pp. 117–134.
- [27] B. Cabral and L. C. Leedom. "Imaging Vector Fields Using Line Integral Convolution". In: *Conference on Computer Graphics and Interactive Techniques. SIGGRAPH '93*. New York, NY, USA: ACM, 1993, pp. 263–270. ISBN: 0-89791-601-8. DOI: [10.1145/166117.166151](https://doi.org/10.1145/166117.166151).
- [28] C. D. Cantrell. *Modern Mathematical Methods for Physicists and Engineers*. New York, NY, USA: Cambridge University Press, 2000. ISBN: 0-521-59827-3.

- [29] H. Carr and D. Duke. "Joint Contour Nets". In: *IEEE Transactions on Visualization and Computer Graphics* 20 (2014), pp. 1100–1113. ISSN: 1077-2626. DOI: [10.1109/TVCG.2013.269](https://doi.org/10.1109/TVCG.2013.269).
- [30] H. Carr and D. J. Duke. "Joint Contour Nets: Computation and properties". In: *IEEE Pacific Visualization Symposium*. Sydney, Australia, 2013, pp. 161–168. DOI: [10.1109/PacificVis.2013.6596141](https://doi.org/10.1109/PacificVis.2013.6596141).
- [31] H. Carr and J. Snoeyink. "Path seeds and flexible isosurfaces using topology for exploratory visualization". In: *Symposium on Data Visualization*. 2003, pp. 49–58. ISBN: 1-58113-698-6.
- [32] H. Carr, J. Snoeyink, and U. Axen. "Computing contour trees in all dimensions". In: *Computational Geometry* 24.2 (2003), pp. 75–94. DOI: [10.1016/S0925-7721\(02\)00093-7](https://doi.org/10.1016/S0925-7721(02)00093-7).
- [33] H. Carr, J. Snoeyink, and M. van de Panne. "Flexible isosurfaces: Simplifying and displaying scalar topology using the contour tree". In: *Computational Geometry* 43.1 (2010). Special Issue on the 14th Annual Fall Workshop, pp. 42–58. ISSN: 0925-7721. DOI: [10.1016/j.comgeo.2006.05.009](https://doi.org/10.1016/j.comgeo.2006.05.009).
- [34] H. Carr et al. "Fiber Surfaces: Generalizing Isosurfaces to Bivariate Data". In: *Eurographics Conference on Visualization (Eurovis)* 34.3 (2015), pp. 241–250. DOI: [10.1111/cgf.12636](https://doi.org/10.1111/cgf.12636).
- [35] W. W.-Y. Chan. "A Survey on Multivariate Data Visualization". In: (2006).
- [36] A. Chattopadhyay et al. "Extracting Jacobi Structures in Reeb Spaces". In: *EuroVis - Short Papers*. Ed. by N. Elmquist, M. Hlawitschka, and J. Kennedy. The Eurographics Association, 2014.
- [37] A. Chattopadhyay et al. "Multivariate topology simplification". In: *Computational Geometry* 58 (2016), pp. 1–24. ISSN: 0925-7721. DOI: [10.1016/j.comgeo.2016.05.006](https://doi.org/10.1016/j.comgeo.2016.05.006).
- [38] Y.-J. Chiang et al. "Simple and optimal output-sensitive construction of contour trees using monotone paths". In: *Computational Geometry* 30.2 (2005), pp. 165–195.
- [39] D. H. Chung et al. "Glyph-Based Multi-field Visualization". In: *Scientific Visualization: Uncertainty, Multifield, Biomedical, and Scalable Visualization*. Ed. by C. D. Hansen et al. London: Springer London, 2014, pp. 129–137. ISBN: 978-1-4471-6497-5. DOI: [10.1007/978-1-4471-6497-5_13](https://doi.org/10.1007/978-1-4471-6497-5_13).
- [40] M. Collins et al. "Chapter 12 - Long-term climate change: Projections, commitments and irreversibility". In: *Climate Change 2013: The Physical Science Basis*. Ed. by IPCC. Cambridge: Cambridge University Press, 2013. DOI: [10.1017/CBO9781107415324.024](https://doi.org/10.1017/CBO9781107415324.024).
- [41] L. Comic and L. D. Floriani. "Dimension-independent simplification and refinement of Morse complexes". In: *Graphical Models* 73.5 (2011), pp. 261–285. DOI: [10.1016/j.gmod.2011.05.001](https://doi.org/10.1016/j.gmod.2011.05.001).
- [42] J. Cox, D. House, and M. Lindell. "Visualizing uncertainty in predicted hurricane tracks". In: *International Journal for Uncertainty Quantification* 3 (2013), pp. 143–156.
- [43] W. De Melo. "Stability and optimization of several functions". In: *Topology* 15 (1976), pp. 1–12.

- [44] G. Debreu. *Theory of Value: An Axiomatic Analysis of Economic Equilibrium*. Cowles Foundation for Research in Economics at Yale University. Monograph. Wiley, 1959. ISBN: 9780300015584.
- [45] C. Deser et al. "Uncertainty in climate change projections: the role of internal variability". In: *Climate Dynamics* 38.3 (2012), pp. 527–546. ISSN: 1432-0894. DOI: [10.1007/s00382-010-0977-x](https://doi.org/10.1007/s00382-010-0977-x).
- [46] Z. Ding et al. "Visual inspection of multivariate volume data based on multi-class noise sampling". In: *The Visual Computer* 32.4 (2016), pp. 465–478. ISSN: 1432-2315. DOI: [10.1007/s00371-015-1070-6](https://doi.org/10.1007/s00371-015-1070-6).
- [47] H. Doleisch, P. Muigg, and H. Hauser. *Interactive Visual Analysis of Hurricane Isabel with SimVis*. <http://www.VRVis.at/>. VRVis Research Center, Vienna, Austria, 2004.
- [48] H. Doraiswamy and V. Natarajan. "Efficient Algorithms for Computing Reeb Graphs". In: *Computational Geometry* 42.6-7 (2009), pp. 606–616. ISSN: 0925-7721. DOI: [10.1016/j.comgeo.2008.12.003](https://doi.org/10.1016/j.comgeo.2008.12.003).
- [49] H. Edelsbrunner and J. Harer. "Persistent homology — a survey". In: *Contemporary Mathematics* 453 (2008), pp. 257–282.
- [50] H. Edelsbrunner, D. Letscher, and A. Zomorodian. "Topological Persistence and Simplification". In: vol. 28. 4. 2002, pp. 511–533. DOI: [10.1007/s00454-002-2885-2](https://doi.org/10.1007/s00454-002-2885-2).
- [51] H. Edelsbrunner and J. Harer. *Computational Topology - an Introduction*. American Mathematical Society, 2010, pp. I–XII, 1–241. ISBN: 978-0-8218-4925-5.
- [52] H. Edelsbrunner and J. Harer. "Jacobi sets of multiple Morse functions". In: *Foundations of Computational Mathematics*. Vol. 312. Cambridge Uni. Press, 2004, pp. 37–57.
- [53] H. Edelsbrunner, J. Harer, and A. K. Patel. "Reeb spaces of piecewise linear mappings". In: *Symposium on Computational Geometry*. 2008, pp. 242–250.
- [54] H. Edelsbrunner, J. Harer, and A. Zomorodian. "Hierarchical Morse Complexes for Piecewise Linear 2-manifolds". In: *Symposium on Computational Geometry*. SCG '01. New York, NY, USA: ACM, 2001, pp. 70–79. ISBN: 1-58113-357-X. DOI: [10.1145/378583.378626](https://doi.org/10.1145/378583.378626).
- [55] H. Edelsbrunner et al. "Local and Global Comparison of Continuous Functions". In: *IEEE Conference on Visualization*. VIS '04. Washington, DC, USA: IEEE Computer Society, 2004, pp. 275–280. ISBN: 0-7803-8788-0. DOI: [10.1109/VISUAL.2004.68](https://doi.org/10.1109/VISUAL.2004.68).
- [56] H. Edelsbrunner et al. "Morse - Smale Complexes for Piecewise Linear 3-manifolds". In: *Symposium on Computational Geometry*. SCG '03. New York, NY, USA: ACM, 2003, pp. 361–370. ISBN: 1-58113-663-3. DOI: [10.1145/777792.777846](https://doi.org/10.1145/777792.777846).
- [57] H. Edelsbrunner et al. "Time-varying Reeb graphs for continuous space-time data". In: *Computational Geometry* 41.3 (2008), pp. 149–166.
- [58] H. Edelsbrunner et al. "Time-varying Reeb Graphs for Continuous Space-time Data". In: *Symposium on Computational Geometry*. SCG '04. New York, NY, USA: ACM, 2004, pp. 366–372. ISBN: 1-58113-885-7. DOI: [10.1145/997817.997872](https://doi.org/10.1145/997817.997872).

- [59] M. Ehrgott and X. Gandibleux. "A survey and annotated bibliography of multiobjective combinatorial optimization". In: *OR-Spektrum - Quantitative Approaches in Management* 22.4 (2000), pp. 425–460. ISSN: 1436-6304. DOI: [10.1007/s002910000046](https://doi.org/10.1007/s002910000046).
- [60] J. Fuchs et al. "A Systematic Review of Experimental Studies on Data Glyphs". In: *IEEE Transactions on Visualization and Computer Graphics* 23.7 (2017), pp. 1863–1879. ISSN: 1077-2626. DOI: [10.1109/TVCG.2016.2549018](https://doi.org/10.1109/TVCG.2016.2549018).
- [61] J. Fuchs et al. "Evaluation of Alternative Glyph Designs for Time Series Data in a Small Multiple Setting". In: *SIGCHI Conference on Human Factors in Computing Systems*. CHI '13. New York, NY, USA: ACM, 2013, pp. 3237–3246. DOI: [10.1145/2470654.2466443](https://doi.org/10.1145/2470654.2466443).
- [62] R. Fuchs, J. Waser, and M. E. Gröller. "Visual Human + Machine Learning". In: *IEEE Transactions on Visualization and Computer Graphics* 15.6 (2009), pp. 1327–1334.
- [63] E. R. Gansner and S. C. North. "An open graph visualization system and its applications to software engineering". In: *SOFTWARE - PRACTICE AND EXPERIENCE* 30.11 (2000), pp. 1203–1233. DOI: [10.1002/1097-024X\(200009\)](https://doi.org/10.1002/1097-024X(200009)).
- [64] C. Garth and X. Tricoche. "Topology - and feature-based flow visualization: Methods and applications". In: *In SIAM Conference on Geometric Design and Computing*. 2005.
- [65] D. Gauld. *Differential Topology: An Introduction*. Dover Books on Mathematics. Dover Publications, 2013, p. 256. ISBN: 9780486319070.
- [66] R. Gautam et al. "Characteristics of meteorological parameters associated with Hurricane Isabel". In: *Geophysical Research Letters* 32.4 (2005). L04801. ISSN: 1944-8007. DOI: [10.1029/2004GL021559](https://doi.org/10.1029/2004GL021559).
- [67] Z. Geng et al. "Visual Analysis of Hurricane Data Using Joint Contour Net". In: *Computer Graphics and Visual Computing (CGVC)*. Ed. by R. Borgo and W. Tang. The Eurographics Association, 2014. ISBN: 978-3-905674-70-5. DOI: [10.2312/cgvc.20141205](https://doi.org/10.2312/cgvc.20141205).
- [68] M. A. Giorgetta et al. "Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5". In: *Journal of Advances in Modeling Earth Systems* 5.3 (2013), pp. 572–597. ISSN: 1942-2466. DOI: [10.1002/jame.20038](https://doi.org/10.1002/jame.20038).
- [69] P. J. Gleckler, K. E. Taylor, and C. Doutriaux. "Performance metrics for climate models". In: *Journal of Geophysical Research: Atmospheres* 113.D6 (2008). ISSN: 2156-2202. DOI: [10.1029/2007JD008972](https://doi.org/10.1029/2007JD008972).
- [70] K. Gruchalla and J. Marbach. "Immersive Visualization of the Hurricane Isabel Dataset". In: 2004.
- [71] D. Günther, J. Salmon, and J. Tierny. "Mandatory Critical Points of 2D Uncertain Scalar Fields". In: *Computer Graphics Forum* 33.3 (2014), pp. 31–40. ISSN: 1467-8659. DOI: [10.1111/cgf.12359](https://doi.org/10.1111/cgf.12359).
- [72] A. Gyulassy and V. Natarajan. "Topology-based simplification for feature extraction from 3D scalar fields". In: *IEEE Transactions on Visualization and Computer Graphics*. 2005, pp. 535–542. DOI: [10.1109/VISUAL.2005.1532839](https://doi.org/10.1109/VISUAL.2005.1532839).
- [73] A. Gyulassy et al. "Conforming Morse-Smale Complexes". In: *IEEE Transactions on Visualization and Computer Graphics* 20.12 (2014), pp. 2595–2603. ISSN: 1077-2626. DOI: [10.1109/TVCG.2014.2346434](https://doi.org/10.1109/TVCG.2014.2346434).

- [74] A. Gyulassy. "Combinatorial Construction of Morse-smale Complexes for Data Analysis and Visualization". AAI3350730. PhD thesis. Davis, CA, USA: University of California at Davis, 2008. ISBN: 978-1-109-06152-9.
- [75] A. Gyulassy, P.-T. Bremer, and V. Pascucci. "Computing Morse-Smale Complexes with Accurate Geometry." In: *IEEE Transactions on Visualization and Computer Graphics* 18.12 (2012), pp. 2014–2022. DOI: [10.1109/TVCG.2012.209](https://doi.org/10.1109/TVCG.2012.209).
- [76] A. Gyulassy et al. "A Practical Approach to Morse-Smale Complex Computation: Scalability and Generality". In: *IEEE Transactions on Visualization and Computer Graphics* 14.6 (2008), pp. 1619–1626.
- [77] A. Gyulassy et al. "Efficient Computation of Morse-Smale Complexes for Three-dimensional Scalar Functions". In: *IEEE Transactions on Visualization and Computer Graphics* 13.6 (2007), pp. 1440–1447. ISSN: 1077-2626. DOI: [10.1109/TVCG.2007.70552](https://doi.org/10.1109/TVCG.2007.70552).
- [78] A. Gyulassy et al. "Topologically Clean Distance Fields". In: *IEEE Transactions on Visualization and Computer Graphics* 13.6 (2007), pp. 1432–1439. ISSN: 1077-2626. DOI: [10.1109/TVCG.2007.70603](https://doi.org/10.1109/TVCG.2007.70603).
- [79] K. Häb et al. "A Data-Driven Approach to Categorize Climatic Microenvironments". In: *Workshop on Visualization in Environmental Sciences (EnvirVis)*. Ed. by K. Rink, A. Middel, and D. Zeckzer. The Eurographics Association, 2016. DOI: [10.2312/envirvis.20161105](https://doi.org/10.2312/envirvis.20161105).
- [80] K. Häb et al. "Spatial Aggregation of Mobile Transect Measurements for the Identification of Climatic Microenvironments". In: *Workshop on Visualization in Environmental Sciences (EnvirVis)*. Ed. by A. Middel, K. Rink, and G. H. Weber. The Eurographics Association, 2015. DOI: [10.2312/envirvis.20151086](https://doi.org/10.2312/envirvis.20151086).
- [81] C. Hansen et al. *Scientific Visualization: Uncertainty, Multifield, Biomedical, and Scalable Visualization*. Mathematics and Visualization. Springer London, 2014. ISBN: 9781447164975.
- [82] W. Harvey and Y. Wang. "Topological Landscape Ensembles for Visualization of Scalar-Valued Functions". In: *Computer Graphics Forum* 29.3 (2010), pp. 993–1002. DOI: [10.1111/j.1467-8659.2009.01706.x](https://doi.org/10.1111/j.1467-8659.2009.01706.x).
- [83] A. Hatcher. *Algebraic topology*. Cambridge: Cambridge University Press, 2002, p. 544.
- [84] C. Heine et al. "A Survey of Topology-based Methods in Visualization". In: *Computer Graphics Forum* 35.3 (2016), pp. 643–667. ISSN: 1467-8659. DOI: [10.1111/cgf.12933](https://doi.org/10.1111/cgf.12933).
- [85] P.-C. Hsu et al. "Investigation of the island-induced ocean vortex train of the Kuroshio Current using satellite imagery". In: *Remote Sensing of Environment* 193 (2017), pp. 54–64. ISSN: 0034-4257. DOI: [10.1016/j.rse.2017.02.025](https://doi.org/10.1016/j.rse.2017.02.025).
- [86] L. Huettenberger, C. Heine, and C. Garth. "Decomposition and Simplification of Multivariate Data using Pareto Sets". In: *IEEE Transactions on Visualization and Computer Graphics* 20.12 (2014), pp. 2684–2693. DOI: [10.1109/TVCG.2014.2346447](https://doi.org/10.1109/TVCG.2014.2346447).

- [87] L. Huettenberger and C. Garth. "A Comparison of Pareto Sets and Jacobi Sets". In: *Topological and Statistical Methods for Complex Data: Tackling Large-Scale, High -Dimensional, and Multivariate Data Spaces*. Ed. by J. Bennett, F. Vivodtzev, and V. Pascucci. Berlin, Heidelberg: Springer Berlin Heidelberg, 2015, pp. 125–141.
- [88] L. Huettenberger, C. Heine, and C. Garth. "A Comparison of Joint Contour Nets and Pareto Sets". In: (2017). Ed. by H. Carr, C. Garth, and T. Weinkauff, pp. 51–65. DOI: [10.1007/978-3-319-44684-4_3](https://doi.org/10.1007/978-3-319-44684-4_3).
- [89] L. Huettenberger et al. *Analyzing Climate Simulation Ensembles Using Pareto Sets*. Poster at the VIS Conference 2017(<http://ieeevis.org/>). Phoenix, AZ, 2017.
- [90] L. Huettenberger et al. "Application of Pareto sets in quality control of series production in car manufacturing". In: *IEEE Pacific Visualization Symposium 00* (2015), pp. 135–139. DOI: [10.1109/PACIFICVIS.2015.7156369](https://doi.org/10.1109/PACIFICVIS.2015.7156369).
- [91] L. Huettenberger et al. "Towards Multifield Scalar Topology Based on Pareto Optimality". In: *Computer Graphics Forum* 32.3 (2013), pp. 341–350.
- [92] L. Hüttenberger et al. *Analysis and visualization of multi-dimensional atmospheric data using Pareto sets*. Talk at the 32nd Conference on Environmental Information Processing Technologies. 2016.
- [93] J.-S. Hwang et al. "Intrusions of the Kuroshio Current in the northern South China Sea affect copepod assemblages of the Luzon Strait". In: *Journal of Experimental Marine Biology and Ecology* 352.1 (2007), pp. 12–27. ISSN: 0022-0981. DOI: [10.1016/j.jembe.2007.06.034](https://doi.org/10.1016/j.jembe.2007.06.034).
- [94] A. Inselberg. "The plane with parallel coordinates". In: *The Visual Computer* 1.2 (1985), pp. 69–91. ISSN: 1432-2315. DOI: [10.1007/BF01898350](https://doi.org/10.1007/BF01898350).
- [95] H. Jänicke et al. "Automatic Detection and Visualization of Distinctive Structures in 3D Unsteady Multi-fields". In: *Eurographics Conference on Visualization (Eurovis)*. EuroVis'08. Chichester, UK: The Eurographs Association; John Wiley; Sons, Ltd., 2008, pp. 767–774. DOI: [10.1111/j.1467-8659.2008.01206.x](https://doi.org/10.1111/j.1467-8659.2008.01206.x).
- [96] H. Jänicke, M. Böttinger, and G. Scheuermann. "Brushing of Attribute Clouds for the Visualization of Multivariate Data." In: *IEEE Transactions on Visualization and Computer Graphics* 14.6 (2008), pp. 1459–1466.
- [97] J. Jeong and F. Hussain. "On the identification of a vortex". In: *Journal of Fluid Mechanics* 285 (1995), pp. 69–94.
- [98] G. Johansson, K. Museth, and H. Carr. "Flexible And Topologically Localized Segmentation." In: *Eurographics Conference on Visualization (Eurovis)*. Ed. by K. Museth, T. Möller, and A. Ynnerman. Eurographics Association, 2007, pp. 179–186. ISBN: 978-3-905673-45-6. DOI: [10.2312/VisSym/EuroVis07/179-186](https://doi.org/10.2312/VisSym/EuroVis07/179-186).
- [99] C. Johnson and C. Hansen. *Visualization Handbook*. Orlando, FL, USA: Academic Press, Inc., 2004. ISBN: 012387582X.
- [100] J. Kehler and H. Hauser. "Visualization and Visual Analysis of Multifaceted Scientific Data: A Survey". In: *IEEE Transactions on Visualization and Computer Graphics* 19.3 (2013), pp. 495–513. DOI: [10.1109/TVCG.2012.110](https://doi.org/10.1109/TVCG.2012.110).

- [101] G. Kindlmann and C. f. Westin. "Diffusion Tensor Visualization with Glyph Packing". In: *IEEE Transactions on Visualization and Computer Graphics* 12.5 (2006), pp. 1329–1336. DOI: [10.1109/TVCG.2006.134](https://doi.org/10.1109/TVCG.2006.134).
- [102] J. Klemelä. "Visualization of Multivariate Density Estimates with Level Set Trees". In: *Journal of Computational and Graphical Statistics* 13.3 (2004), pp. 599–620.
- [103] J. Kniss, G. Kindlmann, and C. D. Hansen. "Interactive Volume Rendering Using Multi-Dimensional Transfer Functions and Direct Manipulation Widgets". In: *IEEE Conference on Visualization*. 2001, pp. 255–262, 562.
- [104] Kronion. <http://www.kronion.de/en/emma/>. [Online; accessed: 19-July-2017].
- [105] A. Kuhn et al. *Topology-Based Analysis for Multimodal Atmospheric Data of Volcano Eruptions*. 2017.
- [106] B. Kuo et al. *Weather Research and Forecasting (WRF) Model simulation data of Hurricane Isabel*. <http://vis.computer.org/vis2004contest/data.html>. Part of the IEEE Visualization 2004 Contest [Online; accessed: 25-August-2017]. 2004.
- [107] O. D. Lampe et al. "Curve-Centric Volume Reformation for Comparative Visualization". In: *IEEE Transactions on Visualization and Computer Graphics* 15.6 (2009), pp. 1235–1242.
- [108] R. Larson and B. Edwards. *Calculus*. Brooks/Cole, Cengage Learning, 2010, p. 1328. ISBN: 9781439030332.
- [109] J. M. Lee. *Introduction to Smooth Manifolds*. Graduate Texts in Mathematics. Springer, 2003. ISBN: 9780387954486.
- [110] J. M. Lee. *Introduction to Topological Manifolds*. Graduate texts in mathematics. Springer, 2000.
- [111] S. Linz et al. "3D fiber probe: State of the art and new developments". In: *Symposium on Optomechatronic Technologies*. 2012, pp. 1–4. DOI: [10.1109/ISOT.2012.6403236](https://doi.org/10.1109/ISOT.2012.6403236).
- [112] R. Liu et al. "Comparative visualization of vector field ensembles based on longest common subsequence". In: *IEEE Pacific Visualization Symposium*. 2016, pp. 96–103. DOI: [10.1109/PACIFICVIS.2016.7465256](https://doi.org/10.1109/PACIFICVIS.2016.7465256).
- [113] S. J. Mascarenhas A. "Implementing time-varying contour trees". In: *Symposium on Computational Geometry* 9 (2005), pp. 370–371.
- [114] J. Mather. *Notes on Topological Stability* John Mather. 1970.
- [115] M. Mihai and R. Westermann. "Visualizing the stability of critical points in uncertain scalar fields". In: *Computers & Graphics* 41 (2014), pp. 13–25. ISSN: 0097-8493. DOI: [10.1016/j.cag.2014.01.007](https://doi.org/10.1016/j.cag.2014.01.007).
- [116] J. Milnor. *Morse Theory*. Annals of mathematics studies. Princeton University Press, 1963. ISBN: 9780691080086.
- [117] M. T. Montgomery et al. "Hurricane Isabel (2003): New Insights into the Physics of Intense Storms. Part I: Mean Vortex Structure and Maximum Intensity Estimates". In: *Bulletin of the American Meteorological Society* 87.10 (2006), pp. 1335–1347. DOI: [10.1175/BAMS-87-10-1335](https://doi.org/10.1175/BAMS-87-10-1335).

- [118] E. Munch and B. Wang. "Convergence between Categorical Representations of Reeb Space and Mapper". In: *Leibniz International Proceedings in Informatics (LIPIcs)* 51 (2016). Ed. by S. Fekete and A. Lubiw, 53:1–53:16. ISSN: 1868-8969. DOI: [10.4230/LIPIcs.SoCG.2016.53](https://doi.org/10.4230/LIPIcs.SoCG.2016.53).
- [119] J. R. Munkres. *Topology*. 2nd ed. Englewood Cliffs, N.J.: Prentice-Hall Inc., 2000, p. 537.
- [120] S. Nagaraj and V. Natarajan. "Relation-Aware Isosurface Extraction in Multifield Data". In: *IEEE Transactions on Visualization and Computer Graphics* 17.2 (2011), pp. 182–191.
- [121] S. Nagaraj and V. Natarajan. "Simplification of Jacobi Sets". In: *Topological Methods in Data Analysis and Visualization*. Ed. by V. Pascucci et al. Mathematics and Visualization. Springer Berlin Heidelberg, 2011, pp. 91–102. ISBN: 978-3-642-15013-5. DOI: [10.1007/978-3-642-15014-2_8](https://doi.org/10.1007/978-3-642-15014-2_8).
- [122] S. Nagaraj, V. Natarajan, and R. S. Nanjundiah. "A Gradient-Based Comparison Measure for Visual analysis of Multifield Data". In: *Computer Graphics Forum* 30.3 (2011), pp. 1101–1110.
- [123] H. A. Nam et al. "Visualizing Nuclear Scission through a Multifield Extension of Topological Analysis". In: *IEEE Transactions on Visualization and Computer Graphics* 18.12 (2012), pp. 2033–2040. ISSN: 1077-2626.
- [124] *National Center for Atmospheric Research*. <https://ncar.ucar.edu/>. [Online; accessed: 07-September-2017].
- [125] M. S. Nixon and A. S. Aguado. *Feature Extraction and Image Processing*. 2nd ed. Orlando, FL, USA: Academic Press, Inc., 2008, p. 88.
- [126] P. Oesterling et al. "Two-stage framework for a topology-based projection and visualization of classified document collections". In: *Visual Analytics Science and Technology*. IEEE, 2010, pp. 91–98.
- [127] V. Pascucci and K. Cole-McLaughlin. "Efficient Computation of the Topology of Level Set". In: *IEEE Conference on Visualization*. 2002, pp. 187–194.
- [128] V. Pascucci et al. "Robust On-line Computation of Reeb Graphs: Simplicity and Speed". In: *ACM Transactions on Graphics* 26.3 (2007). ISSN: 0730-0301. DOI: [10.1145/1276377.1276449](https://doi.org/10.1145/1276377.1276449).
- [129] J. Poco et al. "SimilarityExplorer: A Visual Inter-Comparison Tool for Multifaceted Climate Data". In: *Computer Graphics Forum* 33.3 (2014), pp. 341–350. DOI: [10.1111/cgf.12390](https://doi.org/10.1111/cgf.12390).
- [130] J. Poco et al. "Visual Reconciliation of Alternative Similarity Spaces in Climate Modeling". In: *IEEE Transactions on Visualization and Computer Graphics* 20.12 (2014), pp. 1923–1932. DOI: [10.1109/TVCG.2014.2346755](https://doi.org/10.1109/TVCG.2014.2346755).
- [131] F. H. Post, G. M. Nielson, and G.-P. Bonneau, eds. *Data Visualization: The State of the Art*. Kluwer, 2003. ISBN: 1-4020-7259-7.
- [132] K. Potter et al. "A flexible approach for the statistical visualization of ensemble data". In: *IEEE ICDM workshop on knowledge discovery from climate data*. 2009.
- [133] K. Potter et al. "Ensemble-Vis: A Framework for the Statistical Visualization of Ensemble Data". In: *IEEE International Conference on Data Mining*. Ed. by Y. Saygin et al. Miami, Florida, USA: IEEE Computer Society, 2009, pp. 233–240. DOI: [10.1109/ICDMW.2009.55](https://doi.org/10.1109/ICDMW.2009.55).

- [134] K. Potter et al. "Visualization of uncertainty and ensemble data: Exploration of climate modeling and weather forecast data with integrated ViSUS-CDAT systems". In: *Journal of Physics: Conference Series* 180.1 (2009), p. 012089.
- [135] T. Pryor. "Optics and quality control in the automotive industry". In: *IEEE Journal of Quantum Electronics* 17.12 (1981), pp. 2524–2524. ISSN: 0018-9197. DOI: [10.1109/JQE.1981.1071029](https://doi.org/10.1109/JQE.1981.1071029).
- [136] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2013.
- [137] M. Rautenhaus et al. "Three-dimensional visualization of ensemble weather forecasts - Part 1: The visualization tool Met.3D (version 1.0)". In: *Geoscientific Model Development* 8.7 (2015), pp. 2329–2353. DOI: [10.5194/gmd-8-2329-2015](https://doi.org/10.5194/gmd-8-2329-2015).
- [138] J. Reininghaus and I. Hotz. "Combinatorial 2D Vector Field Topology Extraction and Simplification". In: *Topological Methods in Data Analysis and Visualization*. Ed. by V. Pascucci et al. Springer Verlag, 2011, pp. 103–114. ISBN: 978-3-642-15013-5. DOI: [10.1007/978-3-642-15014-2_9](https://doi.org/10.1007/978-3-642-15014-2_9).
- [139] J. Reininghaus et al. "A Scale Space Based Persistence Measure for Critical Points in 2D Scalar Fields." In: *IEEE Transactions on Visualization and Computer Graphics* 17.12 (2011), pp. 2045–2052.
- [140] O. Rübel et al. "High Performance Multivariate Visual Data Exploration for Extremely Large Data". In: *ACM/IEEE Conference on Supercomputing*. SC '08. Piscataway, NJ, USA: IEEE Press, 2008, 51:1–51:12. ISBN: 978-1-4244-2835-9.
- [141] O. Saeki. "Theory of Singular Fibers and Reeb Spaces for Visualization". In: *Topological Methods in Data Analysis and Visualization IV*. Ed. by T. W. Hamish Carr Christoph Garth. Mathematics and Visualization. Springer International Publishing, 2017, pp. 3–33.
- [142] J. Sanyal et al. "Noodles: A Tool for Visualization of Numerical Weather Model Ensemble Uncertainty". In: *IEEE Transactions on Visualization and Computer Graphics* 16.6 (2010), pp. 1421–1430. DOI: [10.1109/TVCG.2010.181](https://doi.org/10.1109/TVCG.2010.181).
- [143] N. Sauber, H. Theisel, and H.-P. Seidel. "Multifield-Graphs: An Approach to Visualizing Correlations in Multifield Scalar Data". In: *IEEE Transactions on Visualization and Computer Graphics* 12.5 (2006), pp. 917–924.
- [144] R. Scheepens, H. van de Wetering, and J. J. van Wijk. "Non-overlapping Aggregated Multivariate Glyphs for Moving Objects". In: *IEEE Pacific Visualization Symposium*. 2014, pp. 17–24. DOI: [10.1109/PacificVis.2014.13](https://doi.org/10.1109/PacificVis.2014.13).
- [145] D. Schneider et al. "Interactive Comparison of Multifield Scalar Data Based on Largest Contours". In: *Computer-Aided Geometric Design* 30.6 (2013), pp. 521–528.
- [146] D. Schneider et al. "Interactive Comparison of Scalar Fields Based on Largest Contours with Applications to Flow Visualization". In: *IEEE Transactions on Visualization and Computer Graphics* 14.6 (2008), pp. 1475–1482.
- [147] S. Shafii et al. "Topological Features in Glyph-Based Corotation Visualization". In: *Topological Methods in Data Analysis and Visualization III*. Springer International, 2014, pp. 263–276. DOI: [10.1007/978-3-319-04099-8_17](https://doi.org/10.1007/978-3-319-04099-8_17).
- [148] N. Shivashankar and V. Natarajan. "Parallel Computation of 3D Morse-Smale Complexes". In: *Computer Graphics Forum* 31.3pt1 (2012), pp. 965–974. ISSN: 0167-7055. DOI: [10.1111/j.1467-8659.2012.03089.x](https://doi.org/10.1111/j.1467-8659.2012.03089.x).

- [149] N. Shivashankar, M. Senthilnathan, and V. Natarajan. "Parallel Computation of 2D Morse-Smale Complexes". In: *IEEE Transactions on Visualization and Computer Graphics* 18.10 (2012), pp. 1757–1770. ISSN: 1077-2626. DOI: [10.1109/TVCG.2011.284](#).
- [150] S. Smale. "Global analysis and economics - Pareto optimum and a generalization of morse theory". In: *Synthese* 31.2 (1975), pp. 345–358. ISSN: 0039-7857. DOI: [10.1007/BF00485983](#).
- [151] L. Sparling et al. "Case Study on Visualizing Hurricanes Using Illustration-Inspired Techniques". In: *IEEE Transactions on Visualization and Computer Graphics* 15 (2008), pp. 709–718. ISSN: 1077-2626. DOI: [10.1109/TVCG.2008.105](#).
- [152] P. F. Stadler and C. Flamm. "Barrier Trees on Poset-Valued Landscapes". In: *Genetic Programming and Evolvable Machines* 4.1 (2003), pp. 7–20. ISSN: 1389-2576.
- [153] *StereoLithography Interface Specification*. 3D Systems, Inc. 1988.
- [154] A. H. Strahler and A. N. Strahler. *Physische Geographie*. Vol. 3. Ulmer, 2009, p. 688. ISBN: 9783825281595.
- [155] B. Stroustrup. *The C++ Programming Language*. 4th. Addison-Wesley Professional, 2013.
- [156] A. Szymczak. "Stable Morse Decompositions for Piecewise Constant Vector Fields on Surfaces". In: *Computer Graphics Forum* 30.3 (2011), pp. 851–860.
- [157] B. T.F. "Critical points and curvature for embedded polyhedral surfaces". In: *The American Mathematical Monthly* 5 (1970), pp. 475–485.
- [158] R. E. Tarjan. "Efficiency of a Good But Not Linear Set Union Algorithm". In: *ACM Journal* 22.2 (Apr. 1975), pp. 215–225. ISSN: 0004-5411. DOI: [10.1145/321879.321884](#).
- [159] C. Tebaldi and R. Knutti. "The use of the multi-model ensemble in probabilistic climate projections". In: *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 365.1857 (2007), pp. 2053–2075. DOI: [10.1098/rsta.2007.2076](#).
- [160] D. M. Thomas and V. Natarajan. "Detecting Symmetry in Scalar Fields Using Augmented Extremum Graphs". In: *IEEE Transactions on Visualization and Computer Graphics* 19.12 (2013), pp. 2663–2672. ISSN: 1077-2626. DOI: [10.1109/TVCG.2013.148](#).
- [161] J. Tierny and H. Carr. "Jacobi Fiber Surfaces for Bivariate Reeb Space Computation". In: *IEEE Transactions on Visualization and Computer Graphics* 23.1 (2017), pp. 960–969.
- [162] J. Tierny and V. Pascucci. "Generalized Topological Simplification of Scalar Fields on Surfaces." In: *IEEE Transactions on Visualization and Computer Graphics* 18.12 (2012), pp. 2005–2013. DOI: [10.1109/TVCG.2012.228](#).
- [163] X. Tricoche, G. Scheuermann, and H. Hagen. "A Topology Simplification Method for 2D Vector Fields". In: *IEEE Conference on Visualization*. VIS '00. Los Alamitos, CA, USA: IEEE Computer Society Press, 2000, pp. 359–366. ISBN: 1-58113-309-X.
- [164] T. Urness et al. "Strategies for the visualization of multiple 2D vector fields". In: *IEEE Computer Graphics and Applications* 26.4 (2006), pp. 74–82. ISSN: 0272-1716. DOI: [10.1109/MCG.2006.88](#).

- [165] T. Urness and V. Interrante. "Streamline visualization of multiple 2D vector fields". In: *SPIE - The International Society for Optical Engineering*. Vol. 6809. 2008. ISBN: 9780819469816.
- [166] D. M. Ushizima et al. "Augmented Topological Descriptors of Pore Networks for Material Science." In: *IEEE Transactions on Visualization and Computer Graphics* 18.12 (2012), pp. 2041–2050.
- [167] D. J. Valentino, J. C. Mazziotta, and H. K. Huang. "Volume rendering of multimodal images: application to MRI and PET imaging of the human brain". In: *IEEE Transactions on Medical Imaging* 10.4 (1991), pp. 554–562. ISSN: 0278-0062. DOI: [10.1109/42.108590](https://doi.org/10.1109/42.108590).
- [168] J. Vial, J.-L. Dufresne, and S. Bony. "On the interpretation of inter-model spread in CMIP5 climate sensitivity estimates". In: *Climate Dynamics* 41.11 (2013), pp. 3339–3362. ISSN: 1432-0894. DOI: [10.1007/s00382-013-1725-9](https://doi.org/10.1007/s00382-013-1725-9).
- [169] Y. Wan. "Morse theory for two functions". In: *Topology* 14.3 (1975), pp. 217–228. ISSN: 0040-9383. DOI: [10.1016/0040-9383\(75\)90002-6](https://doi.org/10.1016/0040-9383(75)90002-6).
- [170] M. Wand and M. Jones. *Kernel Smoothing*. Chapman & Hall / CRC Monographs on Statistics & Applied Probability. Abingdon, UK: Taylor & Francis, 1994. ISBN: 9780 41255 2700.
- [171] J. Wang et al. "Multi-Resolution Climate Ensemble Parameter Analysis with Nested Parallel Coordinates Plots". In: *IEEE Transactions on Visualization and Computer Graphics* 23.1 (2017), pp. 81–90. DOI: [10.1109/TVCG.2016.2598830](https://doi.org/10.1109/TVCG.2016.2598830).
- [172] G. H. Weber et al. "Feature Tracking Using Reeb Graphs". In: *Topological Methods in Data Analysis and Visualization*. Ed. by V. Pascucci et al. Mathematics and Visualization. LBNL-4226E. Springer Verlag, 2011, pp. 241–253.
- [173] G. H. Weber et al. "Topology-Controlled Volume Rendering". In: *IEEE Transactions on Visualization and Computer Graphics* 13.2 (2007), pp. 330–341.
- [174] G. Weber, P.-T. Bremer, and V. Pascucci. "Topological Landscapes: A Terrain Metaphor for Scientific Data". In: *IEEE Transactions on Visualization and Computer Graphics* 13.6 (2007), pp. 1416–1423. ISSN: 1077-2626. DOI: [10.1109/TVCG.2007.70601](https://doi.org/10.1109/TVCG.2007.70601).
- [175] K. Weiss et al. "A Primal/ Dual Representation for Discrete Morse Complexes on Tetrahedral Meshes". In: *Computer Graphics Forum* 32.3 (2013), pp. 361–370. DOI: [10.1111/cgf.12123](https://doi.org/10.1111/cgf.12123).
- [176] P. Wesseling. *Principles of Computational Fluid Dynamics*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2000. ISBN: 3540678530.
- [177] H. Whitney. "On Singularities of Mappings of Euclidean Spaces. I. Mappings of the Plane into the Plane". In: *Annals of Mathematics* 62.3 (1955), pp. 374–410.
- [178] WikimediaCommons. *File:3D-Leveltorus-Reebgraph.png* — *Wikimedia Commons, the free media repository*. [Online; acc. 26-August-2017]. 2017. URL: <https://commons.wikimedia.org/wiki/File:3D-Leveltorus-Reebgraph.png>.
- [179] WikimediaCommons. *File:Gaussian 2d.svg* — *Wikimedia Commons, the free media repository*. [Online; acc. 26-August-2017]. 2017. URL: https://commons.wikimedia.org/wiki/File:Gaussian_2d.svg.

- [180] WikimediaCommons. *File:Hurricane Isabel from ISS.jpg* — *Wikimedia Commons, the free media repository*. [Online; accessed 26-August-2017]. 2017. URL: https://en.wikipedia.org/wiki/Hurricane_Isabel/media#/File:Hurricane_Isabel_from_ISS.jpg.
- [181] WikimediaCommons. *File:Max_paraboloid.svg* — *Wikimedia Commons, the free media repository*. [Online; accessed 04-September-2017]. 2017. URL: https://commons.wikimedia.org/wiki/File:Max_paraboloid.svg.
- [182] WikimediaCommons. *File:Monkey_saddle_surface.svg* — *Wikimedia Commons, the free media repository*. [Online; acc. 04-September-2017]. 2017. URL: https://en.wikipedia.org/wiki/File:Monkey_saddle_surface.svg.
- [183] WikimediaCommons. *File:Oceanic_gyres.png* — *Wikimedia Commons, the free media repository*. [Online; acc. 30-August-2017] contains materials from the U.S. National Oceanic and Atmospheric Administration. 2015. URL: https://commons.wikimedia.org/wiki/File:Oceanic_gyres.png.
- [184] WikimediaCommons. *File:Saddle_point.svg* — *Wikimedia Commons, the free media repository*. [Online; accessed 04-September-2017]. 2017. URL: https://en.wikipedia.org/wiki/File:Saddle_point.svg.
- [185] A. Wilhelm. *Handbook of Data Visualization*. Ed. by C.-h. Chen, W. Härdle, and A. Unwin. Springer Handbooks Comp. Statistics. Springer, 2008. Chap. II.9.
- [186] A. Wilhelm. “Linked Views for Visual Exploration”. In: *Handbook of Data Visualization*. Ed. by C.-h. Chen, W. Härdle, and A. Unwin. Springer Berlin, 2008, pp. 199–215. DOI: [10.1007/978-3-540-33037-09](https://doi.org/10.1007/978-3-540-33037-09).
- [187] C. Wong, M. C. F. Oliveira, and R. Minghim. “Multidimensional Projections to Explore Time-Varying Multivariate Volume Data”. In: *Conference on Graphics, Patterns and Images*. 2013, pp. 107–114. DOI: [10.1109/SIBGRAPI.2013.24](https://doi.org/10.1109/SIBGRAPI.2013.24).
- [188] J. Woodring and H.-W. Shen. “Multi-variate, Time Varying, and Comparative Visualization with Contextual Cues”. In: *IEEE Transactions on Visualization and Computer Graphics* 12 (2006), pp. 909–916.
- [189] Z. Xie et al. “Exploratory Visualization of Multivariate Data with Variable Quality”. In: *Visual Analytics Science And Technology*. 2006, pp. 183–190. DOI: [10.1109/VAST.2006.261424](https://doi.org/10.1109/VAST.2006.261424).
- [190] E. Zhang et al. “Asymmetric Tensor Analysis for Flow Visualization”. In: *IEEE Transactions on Visualization and Computer Graphics* 15.1 (2009), pp. 106–122. ISSN: 1077-2626. DOI: [10.1109/TVCG.2008.68](https://doi.org/10.1109/TVCG.2008.68).
- [191] F. Zhang, W. Li, and M. E. Mann. “Scale-dependent regional climate predictability over North America inferred from CMIP3 and CMIP5 ensemble simulations”. In: *Advances in Atmospheric Sciences* 33.8 (2016), pp. 905–918. DOI: [10.1007/s00376-016-6013-2](https://doi.org/10.1007/s00376-016-6013-2).
- [192] J. Zhou and M. Takatsuka. “Automatic Transfer Function Generation Using Contour Tree Controlled Residue Flow Model and Color Harmonics”. In: *IEEE Transactions on Visualization and Computer Graphics* 15.6 (2009), pp. 1481–1488. ISSN: 1077-2626. DOI: [10.1109/TVCG.2009.120](https://doi.org/10.1109/TVCG.2009.120).

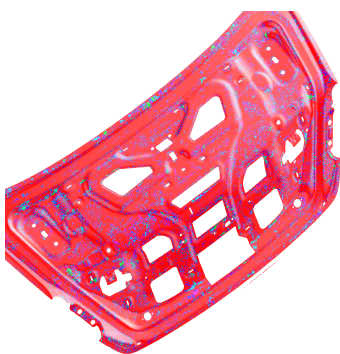
Appendix A

Supplemental materials for Section 7.2

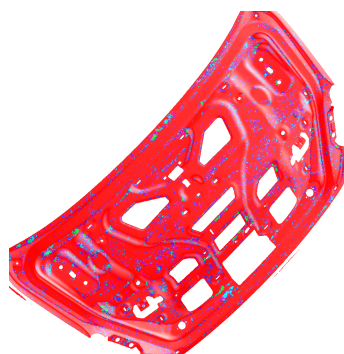
The following images show the weighted Pareto set based on the same distance functions as Figures 7.13 and 7.14. The parameter setting for each image can be retrieved from Table A.1. The weighted Pareto set is colored as described in Section 7.2. Lowest weights (0) and regular areas are colored in red, and the highest weights (6) are colored in green.

ID	Iter.	FWHM	ND	ID	Iter.	FWHM	ND
1	10	6	4	17	20	10	3
2	10	6	5	18	20	10	2
3	20	6	5	19	10	6	3
4	20	10	5	20	10	6	2
5	10	10	5	21	10	3	5
6	10	10	4	22	10	3	4
7	10	10	3	23	10	3	3
8	10	10	2	24	10	3	2
9	20	3	5	25	5	10	5
10	20	3	4	26	5	10	4
11	20	3	3	27	5	10	3
12	20	3	2	28	5	10	2
13	20	6	4	29	5	6	5
14	20	6	3	30	5	6	4
15	20	6	2	31	5	6	3
16	20	10	4	32	5	6	2

TABLE A.1: Parameter settings for the Gaussian smoothing used to create different weighted Pareto set visualizations are indicated by the image ids in the first column.



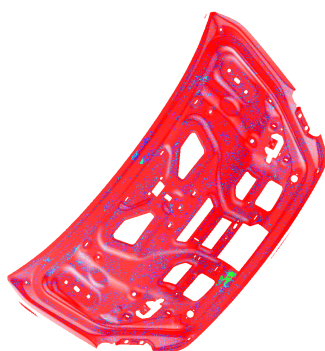
(1)



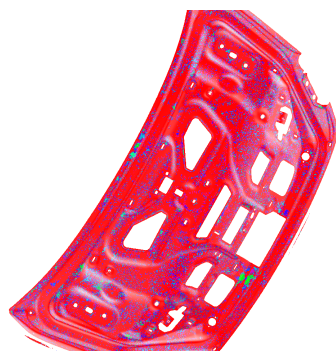
(2)



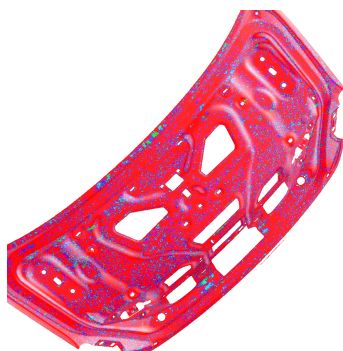
(3)



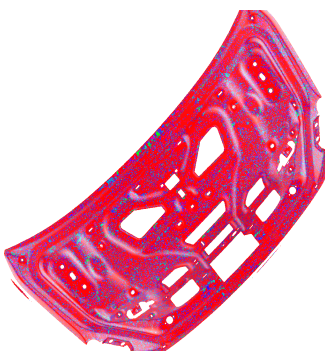
(4)



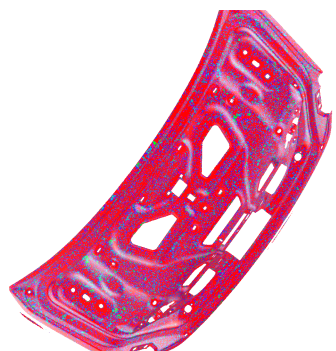
(5)



(6)



(7)



(8)



(9)



(10)



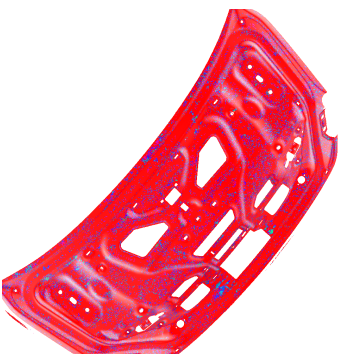
(11)



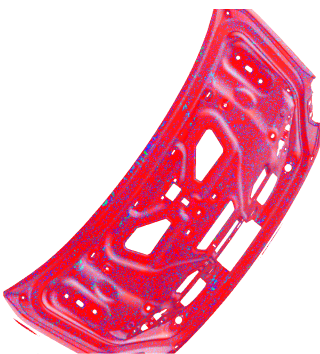
(12)



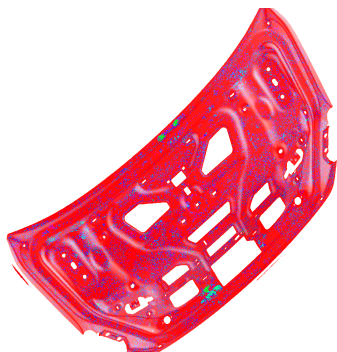
(13)



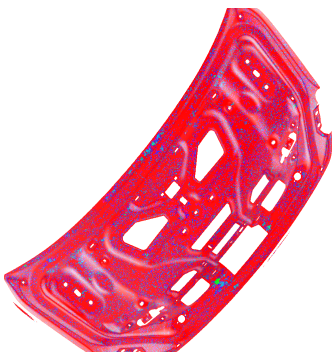
(14)



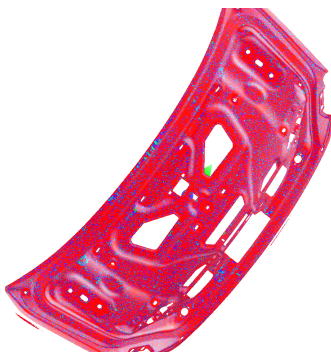
(15)



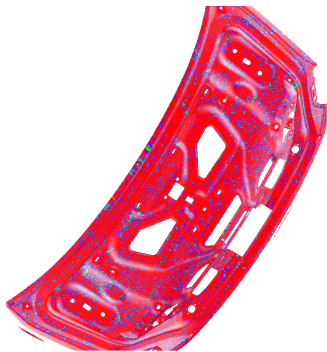
(16)



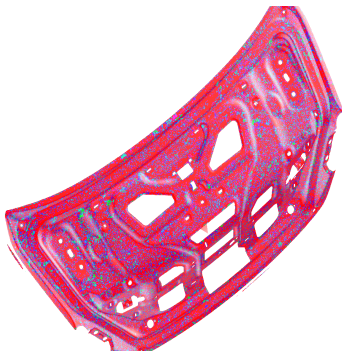
(17)



(18)



(19)



(20)



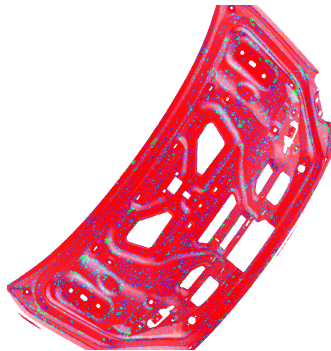
(21)



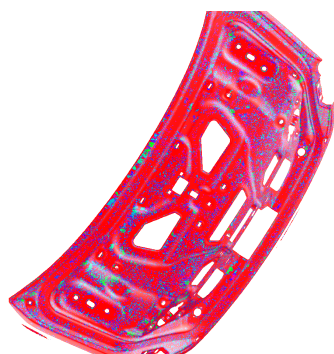
(22)



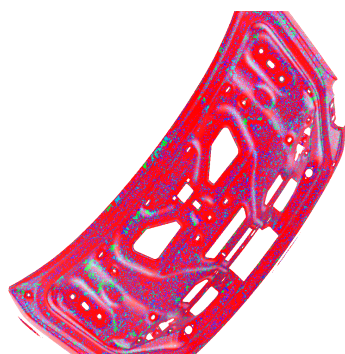
(23)



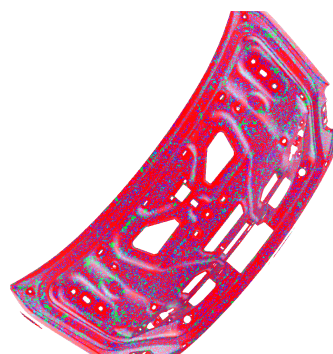
(24)



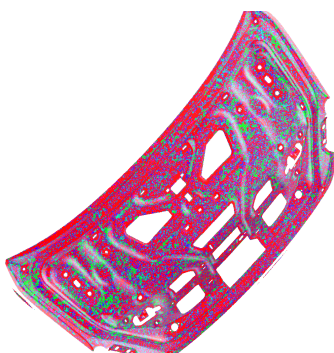
(25)



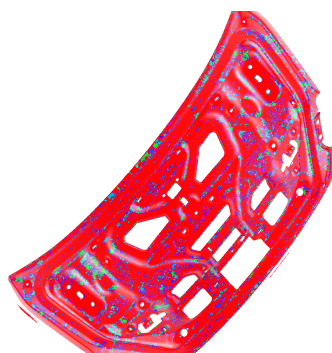
(26)



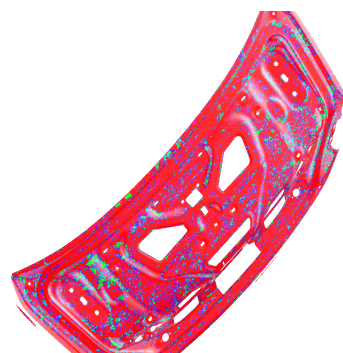
(27)



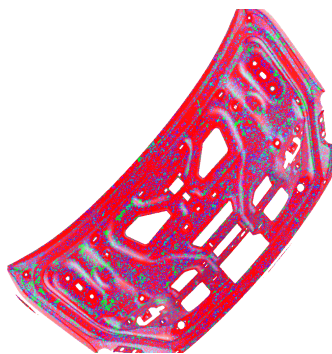
(28)



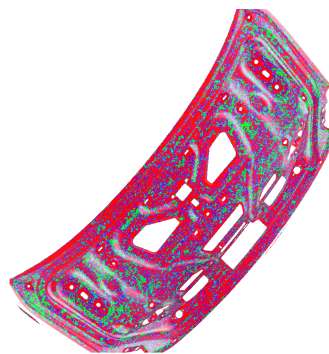
(29)



(30)



(31)



(32)

Appendix B

Curriculum Vitae

B.1 Personal Data

Name: Lars Stefan Hüttenberger
 Date of Birth: April 30, 1986
 Place of Birth: Darmstadt
 Nationality: German

B.2 Education

05/1996 – 05/2005	High school Final grade: 2,8	Jürgen-Fuhlendorf Gymnasium Bad Bramstedt Schleswig-Holstein, GER
10/2006 – 09/2009	Bachelor of Science Major: Computer Science Minor: Mathematics Focus: Algorithmic & Deduction Final grade: 1,9	University of Kaiserslautern Kaiserslautern Rheinland-Pfalz, GER
10/2007 – 09/2015	Research assistant Department of Computer Science & Department of Mathematics	University of Kaiserslautern Kaiserslautern Rheinland-Pfalz, GER
10/2009 – 03/2012	Master of Science Major: Computer Science Minor: Mathematics Focus: Algorithmic & AI Final grade: 1,5	University of Kaiserslautern Kaiserslautern Rheinland-Pfalz, GER
05/2011 – 12/2011	Semester abroad Focus: Machine Learning & Data Mining	University of Regina Regina Saskatchewan, CA

06/2012 – 09/2017	PhD student Focus: Computational Topology & Visualization of Multivariate Data	University of Kaiserslautern Kaiserslautern Rheinland-Pfalz, GER
10/2014 – 10/2014	Visiting student Focus: Topological Aspects of Material Interface Reconstruction	University of California Davis California, USA
07/2015 – 09/2017	Research fellow Department of Computer Science Scientific Visualization Lab	University of Kaiserslautern Kaiserslautern Rheinland-Pfalz, GER

B.3 Work Experience

01/2018 – 10/2018	Project Engineer Quality Assurance Advisory & Services	Commerz Systems GmbH Eschborn Hessen, GER
11/2018 – cont.	Project Engineer Quality Assurance Advisory & Services	Commerzbank AG Frankfurt Hessen, GER

B.4 Publication List

1. **JAguc – a software package for environmental diversity analyses** M. E. Nebel, S. Wild, M. Holzhauser, L. Hüttenberger, R. Reitzig, M. Sperber, T. Stoeck, *Journal of Bioinformatics and Computational Biology* (2011) 9 (6): 749–773
2. **A Survey of Dimension Reduction Methods for High-dimensional Data Analysis and Visualization** D. Engel, L. Hüttenberger, B. Hamann, *Proceedings of IRTG 1131 Workshop 2011* (2012) 27:135–149
3. **Topological Aspects of Material Interface Reconstruction: Challenges and Perspectives** L. Huettenberger, H. Obermaier, C. Garth, K. Joy and H. Hagen, *presented at Topology-based Methods in Visualization 2013*, UC Davis, USA
4. **Towards Multifield Scalar Topology Based on Pareto Optimality** L. Huettenberger, C. Heine, H. Carr, G. Scheuermann, C. Garth, *Computer Graphics Forum* (2013) 32 (3): 341–350
5. **Decomposition and Simplification of Multivariate Data using Pareto Sets** L. Huettenberger, C. Heine, C. Garth, *IEEE TVCG* (2014) 20 (12): 2684–2693
6. **Visualizing the Temporal Development of Thermo-Radiative Features on Ground-based Thermographs** K. Häb, N.H. Feige, L. Huettenberger, A. Middel, A., H. Hagen, *Environmental Earth Sciences* (2014) 72 (10): 3781–3793
7. **A Comparison of Pareto Sets and Jacobi Sets** L. Huettenberger, C. Garth, *Tackling Large-Scale, High-Dimensional, and Multivariate Data Spaces* (2015): 125–141
8. **Application of Pareto Sets in Quality Control of Series Production in Car Manufacturing** L. Huettenberger, N. Feige, A. Ebert, C. Garth, *IEEE Pacific Visualization Symposium*, Hangzhou, China (2015): 135–139
9. **Analysis and Visualization of Multi-Dimensional Atmospheric Data using Pareto Sets** L. Huettenberger, K. Häb, H. Hagen, C. Garth, *presented at 32nd Conference on Environmental Information Processing Technologies* (2016), New Orleans, USA
10. **A Comparison of Joint Contour Nets and Pareto Sets** L. Huettenberger, C. Heine, C. Garth, *Topological Methods in Data Analysis and Visualization IV* (2017): 51–65 -to be published-
11. **The Approximation of Pareto Sets using Directed Joint Contour Nets** J. Bormann, L. Huettenberger, C. Garth, *presented at Topology-based Methods in Visualization 2017*, Tokyo, Japan
12. **Analyzing Climate Simulation Ensembles Using Pareto Sets** L. Huettenberger, K. Feige, M. Böttinger, C. Garth, *Visualization in Practice workshop at the IEEE VIS 2017*, Phoenix, AZ, USA