

A Comprehensive Framework of Tools for Multi-Field Topological Data Analysis

Deepak

Former Assistant Professor of Mathematics

C.R.A. College, Sonipat, Haryana, India

deepakdahiya707@gmail.com

Abstract: *Scientific data often depends on multiple scalar fields. For instance, the study of climate data involves parameters like temperature, pressure, precipitation, and wind speed. The problem of computing distance measures using scalar topology has been studied extensively and proven useful in shape or data clustering, symmetry detection, and feature extraction. However, not all features in a data can be described using the scalar topology. Therefore, developing new tools for computing multi-field or multivariate (consisting of multiple scalar fields) topology is indispensable. Despite its potential, research on building tools in multi-field topology is still in its infancy. In the current paper, we develop data-structures for capturing multi-field topology, design similarity/distance measures for comparing topological features in multi-field data, and demonstrate their effectiveness over measures based on scalar fields.*

Keywords: Multi-field topology, Multivariate data, Scalar fields, Reeb space, Topological data analysis, Topological similarity.

I. INTRODUCTION

Topology is a branch of mathematics which studies the connectivity of shapes. Computational topology, situated at the intersection of mathematics and computer science, is devoted to the development of algorithms to solve topological problems. In recent years, this field has garnered attention for applications in shape and data analysis, leading to the development of a new line of study called Topological Data Analysis. The applications of this discipline span diverse domains such as computational physics, computational chemistry, computational biology, machine learning, computer graphics, and robotics.

This paper aims to develop tools in computational topology for capturing and comparing topological features in shapes or data, with focus particularly on topological analysis based on multiple fields (multi-field or multivariate), in contrast to a single field (scalar). Topological methods are crucial for understanding the spatial behavior and structure of scalar fields in various scientific and engineering problems. Over the last two decades, similarity and distance measures in scalar topology have been developed based on descriptors such as contour tree, Reeb graph, and persistence diagram. The applicability of these measures have been demonstrated in a wide variety of problems such as shape matching [1], classification of bio-molecular or protein structures, noise removal from point-cloud data, symmetry detection, periodicity analysis in time dependant flows, classification of signals, and determining the crystal structure of materials [2].

While scalar topology has been extensively used for single-variable datasets, the growing prevalence of multi-field datasets in scientific applications necessitates the development of efficient tools to analyze and visualize relationships among multiple scalar fields. For example, the nuclear scission point of a Fermium-258 atom has been identified by visualizing the Joint Contour Net (JCN) corresponding to the proton and neutron densities [3]. However, to design similarity measures using multi-field topology one requires to generalize the existing scalar topology based techniques, both theoretically and computationally. Towards this, the focus of the current thesis is on the development of data



structures for capturing multi-field topological features in shapes/data, and formulation of similarity or distance measures between these data structures to facilitate the comparison of shapes/data.

II. RELATED WORK

Similarity and distance measures between scalar fields have been studied extensively using different topological descriptors, such as, contour trees, merge trees, Reeb graphs, extremum graphs and persistence diagrams. A survey paper by Biasotti et al. [4] gives an overview of different topological shape descriptors for scalar fields. Topological shape matching using Reeb graphs and their extensions have been studied extensively by different groups [5, 6]. Hilaga et al. [1] proposed a multi-resolution Reeb graph based similarity measure and applied their method in the shape matching problem. A similar technique was proposed by Zhang et al. [7] to compute a similarity measure between two volumetric scalar data using their multi-resolution dual contour trees and showed an application in molecular shape matching [7]. However, generalizing these works for multi-fields is challenging.

Persistence diagram is one of the most efficient tools for the topological analysis of high-dimensional data. A bottleneck distance was proposed between two persistence diagrams and proven stable. Persistence and bottleneck distance have been widely applied in topological data analysis, noise removal, simplification and image analysis. Dey et al. [8] proposed a persistence distortion metric for comparing metric graphs while Agarwal et al. [9] used the Gromov-Hausdorff distance to compare metric trees. Carlsson et al. [10] proposed the theory of multi-dimensional persistence, although no generalization of the bottleneck distance is known between multi-fields.

Thomas et al. introduced a contour clustering based technique for detecting multiscale symmetry in a volumetric scalar data. Narayanan et al. [11] proposed a topological distance between two extremum graphs for comparing two scalar fields. Various distance metrics have also been proposed based on merge trees. Beketayev et al. [12] proposed a metric between merge trees and showed the effectiveness of their metric by comparing it with the bottleneck distance. Saikia et al. introduced a technique for computing the repeating topological pattern between two scalar data by comparing all subtrees between their merge trees using an extended branch decomposition graph (e-BDG). In another work, Saikia et al. [13] proposed a histogram based distance measure by computing the histograms from the merge trees.

A survey paper by Gao et al. [14] discusses different inexact algorithms in the literature based on the graph edit distance. Fabio et al. proposed an edit distance between Reeb graphs of surfaces and showed the stability property of the proposed metric. Recently, Sridharamurthy et al. presented an edit distance-based metric between two merge trees for feature visualization in a time-varying scalar field data.

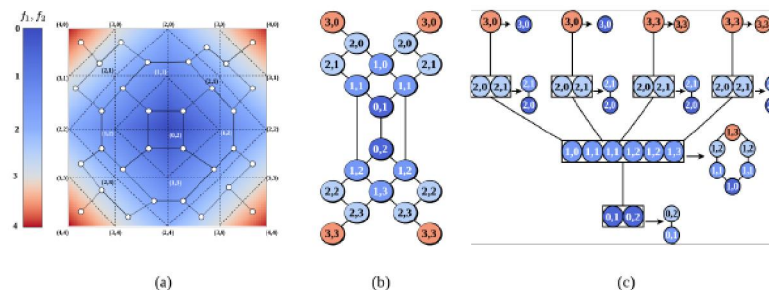


Figure 1 (a) A PL bivariate data over a 2D mesh: The mesh coloring is based on the color map using the first field (f_1), (b) JCN at 4×4 resolution: the nodes are colored based on the values of f_1 , (c) MDRG computed using the algorithm in: the nodes in the first (second)

Various metrics between Reeb graphs such as functional distortion metric and interleaving distance have been proposed. Bauer et al. showed the equivalence dimension are colored based on the values of the first (second) field between the functional distortion metric and the interleaving distance proposed between two Reeb graphs. In a recent paper, Carrière et al. showed that although bottleneck distance between two Reeb graphs is a pseudo-metric, it can



discriminate a Reeb graph from the other Reeb graphs in a small enough neighbourhood, as efficiently as the Gromov-Hausdorff, functional distortion and interleaving distances.

However, research on computing topological similarity measures between multi-fields is still at a nascent stage. Singh et al. [15] proposed a mapper data-structure for capturing the topology of high-dimensional datasets and computed a dissimilarity matrix for the point-cloud data. Dey et al. [16] extended the work to a multiscale mapper (MSM) using a tower of covers and proved its stability by showing the bottleneck distance between the associated persistence diagrams is upper bounded by a distance between the maps. In the current chapter, we propose a multi-resolution Reeb space (MRS) by considering a sequence of Reeb spaces at different resolutions. The sequence of resolutions of MRS can be considered as a special case of the tower of covers in MSM which makes MRS computation more geometric and MRS structure is simplified as a sequence of 1-dimensional simplicial complexes, whereas MSM is a tower of nerves.

Moreover, the bottleneck distance between the persistence diagrams corresponding to two MSMs is a pseudo-metric between the MSMs (similar as the bottleneck distance between Reeb graphs in [6]), whereas we propose a similarity metric between two MRSs. Carr et al. [9] developed a joint contour net (JCN) for a quantized approximation of the Reeb space which is a special case of the mapper structure. Duke et al. showed an effective application of the JCN in visualizing nuclear scission in a multi-field density data. Recently, Agarwal et al. [4] proposed a distance metric between two multi-fields based on their fibre-component distributions and demonstrated its usefulness over scalar-topology. However, a fiber-component distribution is obtained by projecting a Reeb space onto the range, therefore, some features of the Reeb space might be lost. In the current chapter, we propose an improved similarity measure by directly comparing two multi-resolution Reeb spaces.

III. MULTI-DIMENSIONAL REEB GRAPH

A Multi-Dimensional Reeb Graph (MDRG) of a PL multi-field $f: M \rightarrow \mathbb{R}^r$, proposed by Chattopadhyay et al., is a hierarchical decomposition of the Reeb space (or a joint contour net) into a set of Reeb graphs in different dimensions. In particular, to construct the MDRG for a PL bivariate field $f = (f_1, f_2): M \rightarrow \mathbb{R}^2$, first we compute the Reeb graph RG_{f_1} of the field f_1 (in the first dimension). Now each point $p \in RG_{f_1}$ corresponds to a contour, say C_p , of f_1 . Then we restrict function f_2 (in the second dimension) on C_p and define the restricted function as $efp_2 = f_2|_{C_p}$. Finally, we compute the Reeb graphs RG_{efp} for these restricted functions

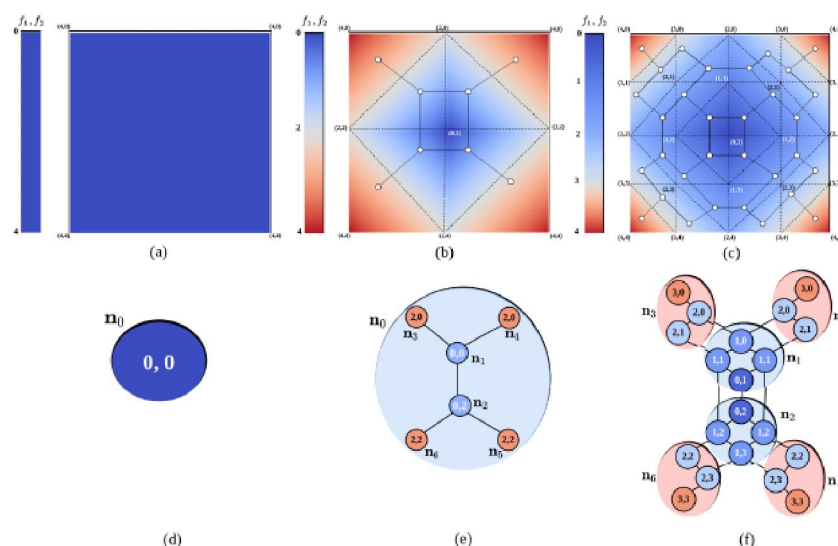


Figure 2: Multi-resolution Reeb Space corresponding to a PL bivariate data: (ring, height). (a) Each component field is quantized into one level, and the corresponding JCN in (d) consists of only one node n_0 . (b) Each component field is quantized into two levels and the corresponding JCN is shown in (e). (c) Each component field is quantized into four levels, and the corresponding JCN is shown in (f). Parent-child relationships between the nodes of JCNs in consecutive Resolutions fp_2 , in the second dimension. Thus MDRG of f , denoted by $MDRG_f$, can be defined as: $MDRG_f = \{(p_1, p_2) : p_1 \in RGf_1, p_2 \in RGf_{p_1}\}$. The definition can be extended for any PL multi-field $f : M \rightarrow R^r$ with $d \geq r \geq 2$. Figure 2(b) shows an example of a quantized Reeb space or JCN for a PL bivariate field (in Figure 2(a)) and Figure 2(c) shows its MDRG. In our algorithm of measuring the similarity between two JCNs, we use the corresponding MDRGs to decide a topologically consistent matching pair of nodes from the JCNs.

IV. CONCLUSION

Topological Data Analysis (TDA) provides a powerful mathematical and computational framework for extracting structural and connectivity information from complex scientific and engineering datasets. While scalar-field topology has been extensively studied using descriptors such as contour trees, Reeb graphs, merge trees, extremum graphs, and persistence diagrams, the increasing availability of multi-field and multivariate data necessitates more expressive topological representations and comparison techniques. This work focuses on computational approaches for capturing, representing, and comparing topological structures in multi-field data. The Multi-Dimensional Reeb Graph (MDRG) provides a hierarchical representation of the Reeb space by decomposing a multi-field into Reeb graphs across successive dimensions. For a bivariate field, the construction starts with the Reeb graph of the first scalar field, followed by restricting the second field to the contours represented by its nodes. The resulting Reeb graphs collectively capture the relationships among multiple scalar fields while preserving important topological characteristics.

Existing similarity measures have demonstrated significant effectiveness in scalar-topology applications, including shape matching, molecular analysis, signal classification, and scientific visualization. However, extending these approaches to multi-field data remains challenging due to higher-dimensional structures and the limited availability of robust distance measures. Therefore, MDRG offers a promising computational foundation for developing topology-aware similarity measures between multi-field datasets. Future work will focus on stable and efficient MDRG distance metrics, higher-dimensional fields, multi-resolution representations, and applications to large-scale scientific and engineering datasets.

REFERENCES

1. Masaki Hilaga, Yoshihisa Shinagawa, Taku Kohmura, and Toshiyasu L. Kunii. Topology Matching for Fully Automatic Similarity Estimation of 3D Shapes. In Proceedings of the 28th annual conference on Computer graphics and interactive techniques, pages 203–212. ACM, 2001. doi:10.1145/383259.383282.
2. Vasileios Maroulas, Cassie Putman Micucci, and Adam Spannaus. A stable cardinality distance for topological classification. Adv. Data Anal. Classif., 14 (3):611–628, sep 2020. ISSN 1862-5347. doi:10.1007/s11634-019-00378-3.
3. David Duke, Hamish Carr, Aaron Knoll, Nicolas Schunck, Hai Nam, and Staszczak. Visualizing nuclear scission through a multifield extension of topological analysis. Visualization and Computer Graphics, IEEE Transactions on, 18, 10 2012. doi:10.1109/TVCG.2012.287.
4. S. Biasotti, L. De Floriani, B. Falcidieno, P. Frosini, D. Giorgi, C. Landi, L. Papaleo, and M. Spagnuolo. Describing shapes by geometrical-topological properties of real functions. ACM Comput. Surv., 40(4), October 2008. ISSN 0360-0300. doi:10.1145/1391729.1391731.
5. Silvia Biasotti, Simone Marini, Michela Mortara, Giuseppe Patan'e, Michela Spagnuolo, and Bianca Falcidieno. 3d shape matching through topological structures. In Ingela Nyström, Gabriella Sanniti di Baja, and Stina Svensson, editors, Discrete Geometry for Computer Imagery, pages 194–203, Berlin, Heidelberg, 2003. Springer Berlin Heidelberg. doi:10.1007/978-3-540-39966-7_18.



6. Silvia Biasotti, Simone Marini, Michela Spagnuolo, and Bianca Falcidieno. Subpart correspondence by structural descriptors of 3d shapes. *Computer-Aided Design*, 38(9):1002 – 1019, 2006. ISSN 0010-4485. doi:10.1016/j.cad.2006.07.003.
7. Xiaoyu Zhang, Marcos, Chandrajit L. Bajaj, and Nathan Baker. Fast matching of volumetric functions using multi-resolution dual contour trees. 2004. URL <https://api.semanticscholar.org/CorpusID:14402435>.
8. Tamal K. Dey, Dayu Shi, and Yusu Wang. Comparing Graphs via Persistence Distortion. In 31st International Symposium on Computational Geometry (SoCG 2015), volume 34 of Leibniz International Proceedings in Informatics (LIPIcs), pages 491–506, Dagstuhl, Germany, 2015. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. ISBN 978-3-939897-83-5. doi:10.4230/LIPIcs.SOCG.2015.491.
9. Pankaj K. Agarwal, Kyle Fox, Abhinandan Nath, Anastasios Sidiropoulos, and Yusu Wang. Computing the gromov-hausdorff distance for metric trees. *ACM Trans. Algorithms*, 14(2), April 2018. ISSN 1549-6325. doi:10.1145/3185466.
10. Gunnar Carlsson and Afra Zomorodian. The theory of multidimensional persistence. *Discrete and Computational Geometry*, 42:71–93, 06 2007. doi:10.1007/s00454-009-9176-0.
11. Vidya Narayanan, Dilip Mathew Thomas, and Vijay Natarajan. Distance between Extremum Graphs. In 2015 IEEE Pacific Visualization Symposium (PacificVis), pages 263–270. IEEE, 2015. doi:10.1109/PACIFICVIS.2015.7156386.
12. Kenes Beketayev, Damir Yeliussizov, Dmitriy Morozov, Gunther H Weber, and Bernd Hamann. Measuring the Distance between Merge Trees. In *Topological Methods in Data Analysis and Visualization III*, pages 151–165. Springer, 2014. doi:10.1007/978-3-319-04099-8_10.
13. Himangshu Saikia, Hans-Peter Seidel, and TinoWeinkauf. Fast similarity search in scalar fields using merging histograms. In *Topological Methods in Data Analysis and Visualization IV : Theory, Algorithms, and Applications*, pages 121–134. Springer, 2015. ISBN 9783319446820. doi:10.1007/978-3-319-44684-4_7.QC 20160406.
14. Xinbo Gao, Bing Xiao, Dacheng Tao, and Xuelong Li. A survey of graph edit distance. *Pattern Analysis and applications*, 13(1):113–129, 2010. doi:10.1007/s10044-008-0141-y.
15. Gurjeet Singh, Facundo Memoli, and Gunnar Carlsson. Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition. In M. Botsch, R. Pajarola, B. Chen, and M. Zwicker, editors, *Eurographics Symposium on Point-Based Graphics*. The Eurographics Association, 2007. ISBN 978-3-905673-51-7. doi:10.2312/SPBG/SPBG07/091-100.
16. Tamal K. Dey, Facundo M'emoli, and Yusu Wang. Multiscale mapper: Topological summarization via codomain covers. In *Proceedings of the Twenty Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA '16*, page 997–1013, USA, 2016. Society for Industrial and Applied Mathematics. ISBN 9781611974331.

