Reeb Space Approximation with Guarantees

Elizabeth Munch
University at Albany – SUNY

Bei Wang
University of Utah

Abstract

The Reeb space, which generalizes the notion of a Reeb graph, is one of the few tools in topological data analysis and visualization suitable for the study of multivariate scientific datasets. First introduced by Edelsbrunner et al. [3], the Reeb space of a multivariate mapping $f : \mathbb{X} \to \mathbb{R}^r$ parameterizes the set of components of preimages of points in $\mathbb{R}^r$. In this paper, we formally prove the convergence between the Reeb space and its discrete approximation, mapper, in terms of an interleaving distance between them. At a fixed resolution of the discretization, such a distance allows us to quantify the approximation quality and leads to a Reeb space approximation algorithm with guarantees based upon established techniques.

1 Reeb Graphs and Reeb Spaces

Multivariate datasets arise in many scientific applications, ranging from oceanography to astrophysics, from chemistry to meteorology, from nuclear engineering to molecular dynamics. Consider, for example, combustion or climate simulation where various physical measurements (e.g. temperature and pressure) or concentrations of multiple chemical species are computed simultaneously. We model these variables mathematically as multiple continuous, real-valued functions defined on a shared domain and represented as a multivariate mapping $f = \{f_1, \cdots, f_r\} : \mathbb{X} \to \mathbb{R}^r$. We are interested in understanding the relationships between these real-valued functions, and more generally, in developing efficient and effective tools for their analysis and visualization.

When $r = 1$, we can study the Reeb graph [6], which contracts each contour (i.e. component of a level set) of a real-valued function to a single point and uses a graph representation to summarize the connections between these contours. When the domain is simply connected, this construction is referred to as a contour tree [8]. Recent work by de Silva et al. [2] has shown that the data of a Reeb graph can be stored in a category-theoretic object called a cosheaf, which opened the way for defining a metric for Reeb graphs known as the interleaving distance.

In the case of multivariate data ($r \geq 1$), we generalize the idea of the Reeb graph to the Reeb space [3]. Let $f : \mathbb{X} \to \mathbb{R}^r$ be a generic, continuous, piecewise linear (PL) mapping defined on a combinatorial d-manifold. Intuitively, the Reeb space of f parametrizes the set of components of preimages of points in $\mathbb{R}^r$ [3]. Two points $x, y \in \mathbb{X}$ are equivalent, denoted by $x \sim y$, if $f(x) = f(y)$ and x and y belong to the same path connected component of the preimage, $f^{-1}(f(x)) = f^{-1}(f(y))$. The *Reeb space* is the quotient space obtained by identifying equivalent points, that is, $\mathbb{RS}_f = \mathbb{X}/\sim$, together with the quotient topology inherited from $\mathbb{X}$. The Reeb graph can then be considered a special case in this context when $r = 1$. Reeb spaces have been shown to have triangulations and canonical stratifications into manifolds [3].

2 JCN and Mapper

Despite its applicability, in practice, a faster algorithm with implementation details is necessary to support the use of Reeb spaces in data analysis. We can approximate the Reeb space with a construction called mapper [7], which takes as input a (potentially multivariate) mapping $f : \mathbb{X} \to \mathbb{R}^r$ and a cover of the range $\mathbb{R}^r$, and produces a summary of the data by constructing the nerve of the connected components of the inverse image of the cover. Such a summary converts the mapping and the covering of the target space into a simplicial complex for efficient computation, manipulation, and exploration [4, 5]. The resulting complex quantizes the variation of multiple variables simultaneously by considering connected components of interval regions (i.e. $f^{-1}(a, b)$ when $r = 1$) instead of the connected components of level sets (i.e. $f^{-1}(c)$). When the mapping is a real-valued function and the covering consists of a collection of open intervals, it is conjectured that mapper converges to the Reeb space as the scale of the covering goes to zero [7], although there are no known proofs to the best of our knowledge.

A special case of mapper for a particular choice of cover is called the Joint Contour Net (JCN), which comes with a simple, efficient algorithm which constructs the nerve of a PL mapping defined over a simplicial mesh involving an arbitrary number of real-valued functions [1]. It can be computed in time

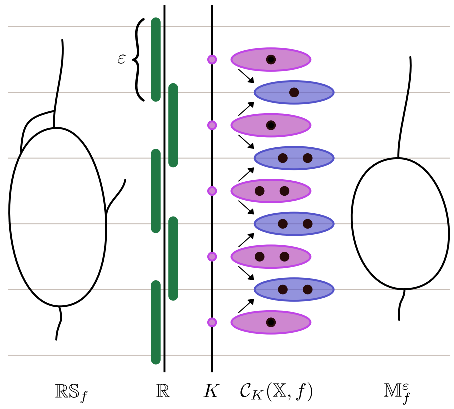

Figure 1: An example of a Reeb space for $r = 1$ (a Reeb graph) with function given by height is shown on the left, denoted as $\mathbb{RS}_f$. Next is the cover for $\mathbb{R}$ given in green, which defines the cell complex K. We can compute the connected components over each element of the cover, shown on the middle, producing a discretization denoted $\mathcal{C}_K(\mathbb{X}, f)$. When this data is turned back into a continuous object, we have the mapper construction on the right, denoted as $\mathbb{M}_f^\varepsilon$. Theorem 2.1, asserts that the interleaving distance between the leftmost and rightmost graphs is bounded by $\epsilon = \mathrm{res}(K)$.

$O(km\alpha(km))$, where m is the size of the input mesh, k is the total number of quantized interval regions, and α is the slow-growing inverse Ackermann function [1]. The authors stated that the JCN can be considered as a discrete approximation that converges in the limit to the Reeb space [1], although this statement was supported only by intuition and, like mapper, lacks any approximation guarantees.

Contributions. In this paper, we formally prove the convergence between the Reeb space and its discrete approximations, JCN and mapper, in terms of the interleaving distance between them. Our work extends and generalizes the tools from the categorical representation of Reeb graphs [2] to a new categorical framework for Reeb spaces. Given a sequence of refined discretization, we prove that these approximations converge to the Reeb space in the interleaving distance.

In particular, we prove that as the resolution of the chosen mesh goes to zero, both JCN and mapper converge to the Reeb space in the interleaving distance. For any set $A \subset \mathbb{R}^n$, define the diameter $\mathrm{diam}(A) = \sup\{\|x - y\| \mid x, y \in A\}$ and let $\mathrm{res}(K) := \sup_{\sigma \in K} \mathrm{diam}(\sigma)$, referred to as the *resolution*. Then we prove the following theorem.

Theorem 2.1 (Reeb Space Approximation)**.** *Given a PL function f defined on a combinatorial d-manifold, let $\mathbb{RS}_f$ represent its Reeb space. Let $\mathbb{M}_f^\varepsilon$ be the mapper output for a particular cover with resolution ε and let $\mathbb{JCN}_f^\varepsilon$ be its joint contour net for a mesh, also with resolution ε. Then*

$$d_I(\mathbb{JCN}_f^\varepsilon, \mathbb{RS}_f) \leq \epsilon$$

and

$$d_I(\mathbb{M}_f^\varepsilon, \mathbb{RS}_f) \leq \epsilon.$$

Based on this theorem, if we construct the JCN or mapper at resolution ϵ, then the distance between the approximate space and $\mathbb{RS}_f$ is at most ϵ – this serves as our approximation guarantee at resolution ϵ. Furthermore, as ϵ goes to zero, the interleaving distance between the approximation and the $\mathbb{RS}_f$ tends to zero, leading to our convergence result.

References

[1] H. Carr and D. Duke. Joint contour nets. *IEEE Transactions on Visualization and Computer Graphics*, 20(8):1100–1113, 2014.

[2] V. de Silva, E. Munch, and A. Patel. Categorification of Reeb graphs. *arXiv*, 2015.

[3] H. Edelsbrunner, J. Harer, and A. K. Patel. Reeb spaces of piecewise linear mappings. In *Proceedings 24th Annual Symposium on Computational Geometry*, pages 242–250, 2008.

[4] P. Y. Lum et. al. Extracting insights from the shape of complex data using topology. *Scientific Reports*, 3, 2013.

[5] M. Nicolaua, A. J. Levineb, and G. Carlsson. Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. *Proceedings National Academy of Sciences of the United States of America*, 108(17):7265–7270, 2011.

[6] G. Reeb. Sur les points singuliers d'une forme de pfaff completement intergrable ou d'une fonction numerique [on the singular points of a complete integral pfaff form or of a numerical function]. *Comptes Rendus Acad.Science Paris*, 222:847–849, 1946.

[7] G. Singh, F. Mémoli, and G. Carlsson. Topological methods for the analysis of high dimensional data sets and 3d object recognition. In *Eurographics Symposium on Point-Based Graphics*, 2007.

[8] M. van Kreveld, R. van Oostrum, C. Bajaj, V. Pascucci, and D. R. Schikore. Contour trees and small seed sets for isosurface traversal. In *Proceedings 13th International Annual Symposium on Computational Geometry*, pages 212–220, New York, 1997. ACM Press.