THE COMPLEX OF TOPOLOGICAL METHODS AND ALGORITHMS APPLICABLE TO DATA ANALYSIS

A.Abzhanov*[1], K.Beketayev[1,2], I.Taimanov[3], Ya.Bazaikin[3], B.Matkarimov[1], M.Ikram[1]

[1]NURIS, Nazarbayev University, Astana, Kazakhstan; *aabzhanov@nu.edu.kz; [2]Lawrence Berkeley National Laboratory, Berkeley, USA; [3]Sobolev Institute of Mathematics, Novosibirsk, Russia

INTRODUCTION.

The main goal of the project is development of new algorithms in the field of topological data analysis and visualization. The abstract presents three new topological methods that allow to incorporate the geometrical information into visualization and representation of topology, and to evaluate the topological error in function approximations.

GEOMETRY-PRESERVING TOPOLOGICAL LANDSCAPES.

We address the problem of enabling comparative capabilities for topological landscapes, and propose a new technique for a direct correlation between a scalar function and its topological landscape. This correlation is accomplished by introducing the notion of geometric proximity into the topological landscapes, reflecting the distance of topological features within the function domain. The proposed technique [1] also makes it possible to compare several complex scalar functions via their topological landscapes, the property that previous versions of the topological landscapes lack.

INTEGRAL PERSISTENCE DIAGRAM.

Classic persistence diagrams present a convenient way to encode the topological information of a function or a space. Each point of the persistence diagram corresponds to a feature and quantifies its importance by absolute difference between the point's two coordinates. A new type of persistence diagrams was developed that has additional stability and can solve high-frequency noise problem [2].

MEASURING ERROR IN APPROXIMATING SUB-LEVEL SET TOPOLOGY.

We address the general problem of quantifying a difference in topology for any two functions, and developed a novel distance definition for topological structures called merge trees. One well-established distance that expresses the topological similarity is the bottleneck distance between persistence diagrams. However, the bottleneck distance does not incorporate sub-level set nesting information, often necessary for analysis. On the other hand, the corresponding merge trees cannot be matched exactly, hinting at a positive difference. To resolve this problem, we developed a novel definition of the distance between merge trees. This distance resembles the bottleneck distance between the persistence diagrams of sub-level sets of the function, but it also respects the nesting relationship between sub-level sets. Furthermore, the proposed distance implicitly distinguishes the noise in the data, similar to the bottleneck distance, resulting in robust measurements resilient to perturbations of the input. This property is crucial when working with scientific data, where noise is a serious obstacle to any analysis.

ACKNOWLEDGMENTS.

The work is supported through the Program 055 "Fundamental and applied scientific research" of the Ministry of Education and Science of the Republic of Kazakhstan.

REFERENCES.

1. K. Beketayev, *et al.* (2012). Geometry-preserving topological landscapes, Proc. WASA'12, Singapore, Nov 26-27, 2012, pp.155-160.
2. A. Abzhanov, Ya. Bazaikin (2014). Stability of integral persistence diagrams, Siberian El. Math. Reports (in press).