KTH Engineering Sciences

Wojciech Chachólski

Stockholm, November 27, 2024

Report on doctor thesis of Niklas Hellmer:
**Probability Meets Topological Data Analysis**

*It is my opinion that the thesis fulfils all the requirements for a doctoral degree in mathematics (stopień doktora w dziedzinie nauk ścistych i przyrodniczych, w dyscyplinie matematyka).*

*Overview of the thesis.* This monograph thesis is based on articles [62], [77], [63], and [91], authored by Niklas Hellmer in collaboration with several co-authors. Additionally, it includes Chapter 6, whose content has not yet been published.

The central theme of this thesis is persistent homology, a fundamental invariant in Topological Data Analysis (TDA). Persistent homology typically takes a distance space as input and produces a persistence diagram as output. Persistent homology is almost never used in static situations. Particularly for applications it is therefore important to study how persistent homology varies when the input changes. To address this dynamic aspect of persistent homology, the thesis employs concepts from probability and measure theory, offering an innovative and intriguing approach to studying the variability of persistent homology.

For instance, persistence diagrams can be interpreted as point measures. This perspective is adopted in Chapter 3, where the concept of the bottleneck profile between persistence diagrams is introduced. These profiles are then used to define a Prokhorov-type distance between persistence diagrams for every admissible function $f\colon [0,\infty[\to [0,\infty[$ (see Definition 3.3.2). The chapter's central result, Theorem 3.3.15, establishes the stability of persistent homology with respect to the Gromov-Hausdorff distance on one side and this newly defined Prokhorov-type distances on the other. In this way, a rich space of distances between persistent diagrams, suitable to study persistent homology, is obtained. In this chapter, Hellmer presents an extensive discussion of the properties of these distances and their associated topology on persistence diagrams. He also illustrates how selecting an appropriate Prokhorov-type distance can enhance the ability to capture and analyze geometrical information more effectively (see Section 3.4).

In my opinion producing relevant distances on outcomes of persistent homology is essential and any result in this direction is valuable. Although, the presented approach to defining distances has proven informative for one-parameter persistent homology, unfortunately it cannot be extended to higher number of parameters.

Chapter 3 is about the range of persistent homology. Chapter 4 is about the domain. Here, Theorem 4.3.1 is the main result which describes exactly the probabilistic information encoded by the limit, in the thermodynamic regime, of the expected Euler characteristic curve (EECC) of the Čech complex (see Definition 2.4.1) of point clouds sampled according to a distribution given by a bounded density. This information is the excess mass transform of the density (see Section 4.1). The statement



KTH Lindstedtsvägen 25,SE-100 44 Stockholm. Phone: +46 8 +46 8 790 7133 . Fax: +46 8 723 1788 . E-mail: wojtek@kth.se
PlusGiro No: 1 56 53-9. Bankgiro No: 895-9223. VAT registration No: SE202100305401. www.math.kth.se

of Theorem 4.3.1 is not the most important part of this chapter however. More important in my opinion is the way it is proven. It relies on the realisation that EECC, for an arbitrary distribution given by a density, can be constructed using the density and the EECC for the uniform distribution (see Theorem 4.2.1 for the formula connecting these EECCs). The key result then follows since EECC for the uniform distribution has a particularly nice form. Theorem 4.2.1 has other interesting consequences for example Corollary 4.2.5, giving an estimation for the $\infty$-norm between EECCs for different distributions.

Chapter 4 is about EECC of Čech complexes and it is quite remarkable, in my opinion, to see such an explicit relation between probabilistic properties of point clouds and certain expected homological properties of their Čech complexes. It would be interesting to understand what are analogous statements for Vietoris-Rips complexes.

The results of Chapter 4 are used in Chapter 5 to test whether a sample from $\mathbb{R}^d$ is drawn from a specific (null) distribution (see Section 5.1.1). For that purpose a procedure called TopoTest is introduced in the thesis. It tests if excess mass transforms of the densities related to the distributions are the same or not. According to Theorem 4.3.1 this is equivalent to testing whether the EECC of the distributions coincides or not. That is a computationally easier task, particularly in dimensions when Euler characteristic of Čech complexes, for example by using alpha complexes, can be calculated effectively. TopoTest relies on Theorem 5.1.4 giving an exponential convergence of certain errors. Several illustrative demonstrations and comparisons of TopoTest are presented. Since EECC of Čech complexes in higher dimensions are not easy to compute, it would be interesting to find a similar test for Vietoris Rips complexes.

The last chapter of the thesis is devoted to generalizations of Dowker construction. For example, a natural filtration on the Dowker complex of a relation is introduced and showed that the transpose homotopy equivalence can be extended to a homotopy equivalence of the filtrations (see Theorem 7.1.4). The most interesting part of this chapter, in my opinion is the introduction of bifiltered probabilistic Dowker complexes in Definition 7.1.6 and proving various stability of this construction (Theorem 7.2.4 for example). This chapter culminates with a low of large number type of theorem 7.2.9 generalising previous results for Vietoris Rips complexes.

The thesis also includes Chapter 6, which focuses on applying Topological Data Analysis (TDA) in signal processing, particularly for detecting vibrations in mechanical systems such as bearings. It is especially rewarding to observe sophisticated mathematical ideas being applied directly to address some practical problems.

*Conclusion.* I find the mathematical content of the thesis novel, important and timely. Part of this work has already been published in reputable journals such as *Discrete & Computational Geometry* and *Statistics and Computing*.

The thesis is well written and beautifully illustrated with informative diagrams and illustrations helping in absorbing presented math. The narrative is well thought through with a generous and informative introduction, providing a valuable resource to other students and the community at large, which I believe is an important aspect of writing a doctoral thesis. In fact I would recommend to all students planning to learn about probabilistic properties of persistent homology to read this thesis.

There are minor misprints scattered throughout the text, such as in the symbols for the measures in Proposition 2.1.7. I recommend that the author carefully reread the manuscript to address these details. While I did not verify every index in all formulas to ensure there are no errors, the arguments and reasoning presented appear sound overall, with one notable exception. In Definition 2.2.11, a shift is defined as an automorphism, yet the key example cited immediately afterward-and referred



KTH Lindstedtsvägen 25,SE-100 44 Stockholm. Phone: +46 8 +46 8 790 7133 . Fax: +46 8 723 1788 . E-mail: wojtek@kth.se
PlusGiro No: 1 56 53-9. Bankgiro No: 895-9223. VAT registration No: SE202100305401. www.math.kth.se

to as the most important example-does not satisfy this requirement. I suggest reconsidering whether the automorphism condition is necessary and possibly eliminating it.

This monograph thesis is based on collaborative work. I am unable to judge the contributions of each of the authors in this join work. I can only base my opinion on the content of the thesis itself and the description of Hellmer's contributions to each of the chapters in the thesis. It is my opinion that the breadth of the thesis, the depth of the presented results, and the quality of writing fulfils all the requirements for a VERY GOOD doctoral thesis in mathematics.

W. Chachólski

Wojciech Chachólski
professor
Department of Mathematics
KTH, 10054 Stockholm
Sweden





KTH Lindstedtsvägen 25,SE-100 44 Stockholm. Phone: +46 8 +46 8 790 7133 . Fax: +46 8 723 1788 . E-mail: wojtek@kth.se
PlusGiro No: 1 56 53-9. Bankgiro No: 895-9223. VAT registration No: SE202100305401. www.math.kth.se