



# GDAŃSK UNIVERSITY OF TECHNOLOGY

FACULTY OF APPLIED PHYSICS AND MATHEMATICS

---

Gdańsk, December 2, 2024

Dr. hab. Paweł Pilarczyk, prof. of Gdańsk Tech  
Division of Differential Equations  
and Applications of Mathematics  
Institute of Applied Mathematics  
Faculty of Applied Physics and Mathematics  
Gdańsk University of Technology  
G. Narutowicza 11/12  
80-233 Gdańsk, Poland

## REVIEW

### of the PhD dissertation of Niklas Hellmer entitled “Probability Meets Topological Data Analysis”

In accordance with paragraphs 1 and 2 of Article 187 of the Act of 20 July 2018 “Prawo o szkolnictwie wyższym i nauce” (as published in Dz. U. 2023 at position 742, with later amendments), a doctoral dissertation in the field of exact and natural sciences in the discipline of mathematics should reflect the candidate’s general theoretical knowledge and the ability to conduct scientific work; moreover, the subject of the doctoral dissertation should be an original solution to a scientific problem. I am firmly convinced that the doctoral dissertation of Mr. Niklas Hellmer entitled “Probability Meets Topological Data Analysis” that is the subject of this review **meets the requirements for doctoral dissertations** set forth by this act. In order to justify this, let me discuss the submitted doctoral dissertation.

The doctoral dissertation of Mr. Niklas Hellmer, supervised by Dr. hab. Paweł Dłotko, contains thorough description of several inter-related results at the intersection of topological data analysis and statistics. Most results are co-authored with other scientists, and

contribution of the co-authors is clearly stated at the beginning of each chapter. It follows that the contribution of the candidate has been considerable, which justifies including these results in his doctoral dissertation.

The dissertation is written in English. Its structure is well thought out and helps following the achievements of the candidate for the PhD degree. The Abstract in English and in Polish as well as the Introduction section provide concise and adequate overview of the contents of the dissertation. The Background section introduces several definitions and information necessary to appreciate the material contained in the core chapters of the dissertation, and makes the dissertation self-contained.

The description of the results achieved by Mr. Hellmer are split into five sections. The first result is the introduction of bottleneck profiles as a generalization of the bottleneck distance used as a metric on the space of persistence diagrams. Their relation to Wasserstein distance (another metric for persistence diagrams) is then established, and discrete Prokhorov metrics for persistence diagrams are introduced. Thorough analysis of various properties of the introduced notions is supported by numerical experiments that are carefully designed, conducted and illustrated. This helps understanding the notions introduced and convinces about their usefulness.

The next section introduces a theorem saying that under certain assumptions, the expected Euler characteristic curves of the Čech complex of a finite number of points sampled with respect to two distributions are the same if and only if their excess mass transforms agree. It is worth pointing out is that the research on this topic includes constructive proofs that help using these results for practical applications.

Another section is devoted to designing goodness of fit tests based on the Euler characteristic curve of a sample. In addition to proving some theoretical properties of the tests that justify their usefulness, algorithms for the one-sample and two-sample testing are provided and explained in detail. The most spectacular feature of this method to me is that it works in arbitrary dimensions, unlike most “classical” statistical tests, but this should have actually been expected, because this method is based on topology. The ideas behind the tests and their features are illustrated by several examples and numerical experiments that include a thorough comparison against the Kolmogorov-Smirnov test on a multitude of distributions. This detailed discussion is justified by the novelty of the method introduced and shows its promising capabilities.

Section 6 concerns an application of topological data analysis methods to the detection of broken bearings on the basis of signal (sound) recorded while they work. The authors

propose an interesting method. Its superiority over state-of-the art methods is convincingly explained and evidence is shown with numerical simulations. A laboratory test was conducted with some real bearings, and additionally sound samples were taken from a bearing installed in an idler in a real belt conveyor at some industrial facility. However, it seems that the actual amount and diversity of data collected was not satisfactory for a reliable conclusion of this experiment. I must admit that the analysis of these real signals is the least convincing achievement of the dissertation to me, even though the resulting accuracy was very high. As I understood, the classification was conducted on segments cut out from very few long input signals, which might have resulted in repeated patterns appearing in the data samples. I perceive this a flaw in experimental design.

The last chapter of the dissertation concerns – roughly speaking – a new approach to persistence of a point cloud that overcomes two very annoying features of the “classical” persistent homology of a point cloud: sensitivity to outliers and disregarding local density of the data points. This approach is based on a density-sensitive bifiltration of Dowker complexes. I perceive the new approach very valuable and of great potential for applications. I would like to encourage Mr. Hellmer to further pursue his research on this topic, to promote this approach, and to contribute towards the development of user friendly software that would not only allow one to apply this method to real data, but would also help understand the results obtained.

In general, I feel that the author of the PhD dissertation under review has proved his ability to use advanced mathematical research methods. One can see skills in precisely formulating mathematical definitions, theorems, and proofs. The comprehensive bibliography and referring to a variety of recent publications prove good knowledge of the state of the art and current research on the topic. I feel that the entire doctoral dissertation presents the general theoretical knowledge at the level corresponding to the doctoral degree in the discipline of mathematics, and shows the ability to conduct scientific work at professional level. The mathematical notation used throughout the dissertation is appropriate and consistent. The dissertation contains original mathematical results that contribute towards solving specific mathematical problems. After each chapter, possible directions of future research are outlined, which shows that the candidate is an active researcher and has some research plans for the future.

The dissertation is written in a clear and careful way, in very high quality English. However, the Polish version of Abstract contains some slips, mainly regarding the typography (wrong hyphenation, incorrect quotation marks, a grammar mistake “kompleksu Čech” in

the middle, short words like “i” left at the end of lines). In the entire text, I was only able to find mistakes or faults of minor importance; let me list most them:

- page v, line –11: “relaibly” instead of “reliably” (a typo)
- page 2, line 21: “The bottleneck profiles arises” should be: “... arise”
- page 7, Definition 2.1.2: “Let  $(X, d)$  a metric space” should be: “Let  $(X, d)$  be a metric space” (“be” is missing)
- page 7, line –2: the word “Borel” should not be there, because the definition of a general  $\Sigma$ -algebra follows
- page 12, Definition 2.2.2: the non-empty sets should be assumed finite, or otherwise Definition 2.2.3 might not work
- page 13, line –9: “continnuous” (double ‘n’)
- page 22: before section 2.2.3,  $m$  denotes a number between 0 and 1, while in section 2.2.3 it is a natural number (see e.g. Figure 2.5 on page 23); this is misleading
- page 25, line 6: “hence it we state it” (the first “it” is superfluous)
- page 30, line –13: “the” is repeated in “over the the faces”
- page 32, Definition 2.3.10: “is multiset of points”  $\rightarrow$  “is a multiset of points” (the article “a” is missing)
- page 33, line –2:  $\eta(X)$  in the formula should actually be  $\eta(x)$
- page 35, Definition 2.3.16: “the empty diagram”: note that every persistence diagram contains the diagonal, see Definition 2.3.10, so calling a diagram without off-diagonal points empty and denoting it with the empty set symbol is imprecise
- page 41, line –3: “an discussion” should be: “a discussion”
- page 42, line –11: “and one in the right”: actually, there are two bottlenecks in the right panel of Figure 3.1 (the information in the caption of the figure is correct)
- page 58, line –10: “of it alpha complex” should be: “of its alpha complex”
- page 59, Figure 3.7: the “1 vs 2” bars are hard to see; a good idea might be to shift the three bars slightly to the sides (in the horizontal direction) to make them easily distinguishable
- page 59, Figure 3.9: a color bar is necessary to show the correspondence between the colors and the values they represent
- page 60, in the first two lines of the “Parameter Tuning” paragraph: “The set off all such functions”: should be “of” instead of “off”
- page 65, line –7: “numpy” should be: “NumPy” (see <https://numpy.org/>)
- page 69, line –3: “amient” should be: “ambient”
- page 83, line 4: “Using the Lemma 5.1.2 we obtain following theorem”: the article “the” should be moved: “Using Lemma 5.1.2 we obtain the following theorem”
- page 85, line –14: “in later” should be: “in the latter”
- page 87, line 1: “although some some asymptotic” (the word “some” appears twice)
- page 91, line 2 of section 5.2.2: “by the Algorithm 5.2” (the article “the” shouldn’t be there); a few more problems with the article “the” appear, but I am not listing them
- page 108, line –8: “followingthe” (the space is missing)

- page 110, in the exposed formulas for  $r_1$  and  $r_k$ , in the last index there should be  $(d - 1)\tau$  instead of  $d\tau$
- page 110, lines  $-3$  and  $-4$ : “dim” should be replaced with “ $d$ ”; the same on page 122, line  $-9$
- page 111, line  $-8$ : “we would it regard it”: the first “it” should be deleted
- page 128, line 1: “weigh” should be: “weight”
- page 136, line  $-12$ : “such no more” should be: “such that no more”
- page 141, the first line after the exposed formula for  $\Lambda_{ij}$ : “innstance” (double ‘n’)

It is important to add that the PhD dissertation is accompanied by actual software written by the candidate or to which the candidate has contributed. This is mainly an open source implementation of the algorithms introduced in the dissertation, such as computing the bifiltered measure Dowker complex, or code to reproduce the experiments described in the dissertation. I consider these additions valuable. They also show programming skills of the candidate.

Last but not least, let me mention that Mr. Niklas Hellmer is a co-author of four academic papers: one published in *Statistics and Computing* (2023 Impact Factor 1.6, Polish Ministry 140 points), another one published in *Discrete & Computational Geometry* (2023 Impact Factor 0.6, Polish Ministry 100 points), two arXiv preprints (one of them submitted to *Advances in Applied Mathematics* with 2023 Impact Factor 1.0 and Polish Ministry evaluation of 100 points), and a database publication in *Scientific Reports* (2023 Impact Factor 3.8, Polish Ministry 140 points).

**Summary.** After having reviewed the doctoral dissertation of Mr. Niklas Hellmer entitled “Probability Meets Topological Data Analysis,” I am convinced that it meets the requirements set forth for doctoral dissertations by the Act mentioned at the beginning of the review. I believe that the dissertation is an original solution to a scientific problem, and contains original mathematical methods and valuable scientific results. The presented dissertation indicates that its author has general theoretical knowledge expected from a doctoral degree holder in the field of exact and natural sciences in the discipline of mathematics, and has mastered the research methods necessary to conduct scientific work. I therefore ask for admitting Mr. Niklas Hellmer to the further stages of the procedure for awarding the PhD degree in the discipline of mathematics. Moreover, I propose that this doctoral dissertation be distinguished, because I believe that the results achieved have significant scientific importance, and the dissertation itself has been written in a very thorough and careful way.