Skoltech
Skolkovo Institute of Science and Technology

Jury Member Report – Doctor of Philosophy thesis.

Name of Candidate: Georgii Novikov

PhD Program: Computational and Data Science and Engineering

Title of Thesis: Tensor-train decomposition for efficient data representation and learning

Supervisor: Professor Ivan Oseledets
Co-supervisor: Assistant Professor Maxim Panov

Name of the Reviewer: Ivan Tyukin

I confirm the absence of any conflict of interest (Alternatively, Reviewer can formulate a possible conflict)	**Date: 04-12-2024**

The purpose of this report is to obtain an independent review from the members of PhD defense Jury before the thesis defense. The members of PhD defense Jury are asked to submit signed copy of the report at least 30 days prior the thesis defense. The Reviewers are asked to bring a copy of the completed report to the thesis defense and to discuss the contents of each report with each other before the thesis defense.

If the reviewers have any queries about the thesis which they wish to raise in advance, please contact the Chair of the Jury.

Reviewer's Report

Reviewers report should contain the following items:

- Brief evaluation of the thesis quality and overall structure of the dissertation.
- The relevance of the topic of dissertation work to its actual content
- The relevance of the methods used in the dissertation
- The scientific significance of the results obtained and their compliance with the international level and current state of the art
- The relevance of the obtained results to applications (if applicable)
- The quality of publications
- The summary of issues to be addressed before/during the thesis defense

The thesis focuses on the development of a broad family of machine learning algorithms exploiting the advantage of low-rank tensor-train decomposition. The overall structure of the thesis is aligned with the four general directions: density estimation, compression of linear layers, compression of activations (few-bit backward), and data representation (tensor-train point cloud). These correspond to four chapters in the thesis (chapters 2 – 5) describing each method in more detail.

The combination of these broad topics constitutes a significant contribution to the area of learning and data representation and is in full agreement with the title of the thesis. The methods used in the thesis range from numerical and computational simulations to tools from numerical linear algebra and analysis. These are all relevant and appropriate for the spectrum of problems considered in these works.

Scientific significance of results and their international standing in relation to state-of-the-art in the area are supported by their publication in extremely competitive international venues. I have no doubts that the material presented in the thesis could justify the award of a PhD.

Having said this, and upon careful reading through the thesis, I have few minor queries which I am sure the author will be able to address. Below I provide a non-exhaustive list of examples of these small issues which could be taken as a guidance whilst preparing the final submission.

1. Structure. The thesis reads as a compilation of published work, with very clearly noticeable gaps between chapters. Moreover, the author uses “work” when referring to individual chapters which leaves an impression of a hastily job. Please consider rewriting and put effort in presenting the material as a thesis “glued” together by a consistent and coherent story. Conclusion is extremely short and does not contain any vision for “what’s next”. In the current version, it’s just a list of 4 topics considered in the thesis and nothing more.
2. Language and presentation. The language is too sloppy at times. I recommend a thorough overhaul. The same applies to few notational collisions (please see below). In the beginning of Chapter 1 there is a “hanging” diagram which is not referred to anywhere in the text (at least I did not find a ref). Please correct.
3. The choice of notation (indices) is inconsistent with the one used to define the Kronecker symbol (page 22) and the Identity matrix (same page)
4. Definition 1.2.1 is not a definition. Please rephrase
5. The term “good approximation” is not defined (page 31). It is not clear in which sense an approximation could be deemed good and why.
6. I don’t think that “reshape” was ever defined in the text (page 31)
7. Statements (formal statements highlighted in blue). Please add references to appropriate papers and monographs these statements are taken from. Otherwise present proofs in Appendix. The same (modulo proofs) applies to Definitions
8. Page 35. The example explaining the computational/memory efficiency of tensor train representations needs a better presentation and explosion. This property is central for all subsequent chapters and I would expect that the author could put more effort in making this point crystal-clear. Please add more examples and explanations
9. Page 37. Please define a quadratic function defined on tensors
10. Page 39. Please explain why computing the gradient is needed without computing the vector of partial derivatives first. It is clear that there could be additional costs arising within the naive approach. Please explain which costs, and give examples
11. Please reformat eq in the statement (1.44)

12. Please explain what do you mean when referring to the “curse of dimensionality”. Why and how it is being overcome (in refs 17-20)? This will enable you to highlight better the difference between your approach and theirs.
13. Page 48. I don’t understand (2.61). Please define what approx means here
14. The text around (2.70). The proposal is to work with tensor train format of the coefficients. Are there any assumptions which could justify the existence of decomposition/representations enabling to describe the object of interest with small number of variables?
15. Page 52, (2.75). x_i is bold. Isn’t it a scalar (and hence shouldn’t be in bold)?
16. Page 55. Please justify why “negative” values of densities could be neglected. The current explanation is too vague and loose.
17. Page 57. Sometimes the function q_\theta has one argument, sometimes two, and sometimes k. Please add clarifications as appropriate
18. Page 58. The order of symbols in the last line of the expression inside the integral is inconsistent with notational agreements used in the thesis
19. Page 64. Please explain what do you mean by saying that an assumption is “temporary”. An assumption is either there or not. Please clarify
20. In chapters 3-5, it would be helpful if some analysis or numerical stability of Algorithms could be given

As stated before, the list is not exhaustive but is indicative enough of the minor changes which could be made to the final version of the thesis.

All in all, the above technical issues do not impact my overall very positive opinion about this work. I believe that the thesis can recommended for the formal defense.

/Provisional Recommendation

■ *I recommend that the candidate should defend the thesis by means of a formal thesis defense*

☐ *I recommend that the candidate should defend the thesis by means of a formal thesis defense only after appropriate changes would be introduced in candidate’s thesis according to the recommendations of the present report*

☐ *The thesis is not acceptable and I recommend that the candidate be exempt from the formal thesis defense*