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: Alexander Kulikov

I confirm the absence of any conflict of interest	**Date: 15-11-2024**

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

- Brief evaluation of the thesis quality and overall structure of the dissertation.

The thesis, *"Tensor-Train Decomposition for Efficient Data Representation and Learning,"* presents an exploration of low-rank tensor-train decomposition applied to various aspects of deep learning, such as density estimation, memory reduction in neural networks, and compression of large high-dimensional point clouds. The work is well-structured, organized into chapters, each dedicated to a specific application of tensor-train decomposition. Each chapter follows a logical sequence, containing an introduction, necessary background information, methodology, experiments, and conclusions. The first chapter provides an introduction and establishes the necessary mathematical notations, definitions, theorems, and algorithms needed to understand the main body of the text. The quality of writing and presentation aligns well with academic standards.

- The relevance of the topic of dissertation work to its actual content

The topic of applying tensor decompositions in deep learning is both important and relevant, addressing modern challenges in this field. The study's content is fully aligned with its research topic and advances related issues in efficient data representation, learning optimization, and new deep learning methods.

- The relevance of the methods used in the dissertation

The methods used in the thesis are well-chosen and relevant: modern tensor decomposition techniques are applied, a wide range of new learning algorithms are utilized, and advanced numerical linear algebra methods are appropriately employed, which is necessary for the effective implementation of the approaches presented in the paper.

- The scientific significance of the results obtained and their compliance with the international level and current state of the art

This work makes a significant scientific contribution to the field. The developed methods for applying tensor-train decomposition to density estimation, neural network memory reduction, and point cloud compression are in line with the international state-of-the-art and offer innovative solutions that improve computational efficiency. New, more efficient neural network training schemes address a pressing modern problem, and the presented memory-saving backpropagation schemes are particularly important in practice for the further scaling of neural network architectures. The developed density estimation method enables a broader class of tractable probabilistic inference queries, which is relevant for machine learning domains such as causal machine learning. Altogether, this work makes a meaningful contribution to computational science and engineering. The scientific results demonstrate a high level of rigor and originality, consistent with international standards.

- The relevance of the obtained results to applications (if applicable)

The methods developed in this thesis have broad practical applications, especially given the developed publicly available open-source packages (e.g., TTDE, Few-bit Backward, TTMLinear). These packages facilitate further exploration and application of TT-based methods in real-world machine learning models. The applications to GPT models and point cloud

compression emphasize the utility of the developed methods in both commercial and academic contexts, affirming the practical relevance of the work.

- The quality of publications

The publications associated with this research have been presented at respected international machine learning conferences, notably those with top-tier A and A* scores, including Uncertainty in Artificial Intelligence (UAI) 2021 and the International Conference on Machine Learning (ICML) 2023, as well as several other conferences, such as the Pacific Asia Conference on Language, Information, and Computation (PACLIC) 2023 and the International Conference on Computational Optimization (ICOMP) 2024. This publication record reflects a high quality of research and indicates that the work has undergone peer review, which attests to its scientific relevance and acceptance within the broader research community.

- The summary of issues to be addressed before/during the thesis defense

The thesis is well-written and fully ready for defense as is. However, the following remarks could be considered to further enhance the work's qualification:

1. The topic of future work is not fully addressed. Adding reflections on the challenges and future directions for the developed methods in the conclusions of the chapters could enrich the work.
2. Comparison with alternative decomposition techniques: A comparative discussion of TT decomposition with alternative low-rank tensor methods would be beneficial. It is not always clear why tensor-train decomposition is chosen over alternative decomposition methods.

Provisional Recommendation

x *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*