Skoltech
Skolkovo Institute of Science and Technology

Jury Member Report – Doctor of Philosophy thesis.

Name of Candidate: Talgat Daulbaev

PhD Program: Computational and Data Science and Engineering

Title of Thesis: Applications of differential equations and reduced-order modeling for deep learning

Supervisor: Professor Ivan Oseledets
Co-supervisor: Professor Andrzej Cichocki

Name of the Reviewer:

I confirm the absence of any conflict of interest (Alternatively, Reviewer can formulate a possible conflict)	**Date: 06-03-2023**

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

This thesis provides a series of research works on neural ODEs and deep networks. For neural ODEs, it well covers the techniques on acceleration, normalization, and adversarial robustness. For deep networks, sketching is taken for a reduced-order network modeling, and active subspace is analyzed for network compression and adversarial attack. Generally, the topic of dissertation works well to the actual content. The scientific contribution is enough as a good PhD dissertation among the current state-of-the-art. The quality of most publications is very good.

Before the thesis defense, perhaps there are some issues to be addressed as follows:

1. It may be more friendly for readers if the implementation details of a neural ODE for the 1st example in chapter 2, such as the reason for catastrophic instabilities, the network architecture.
2. The work may be more attractive if some real-data based experiment for tasks such as probability density estimation.
3. Can you explain more why barycentric Lagrange interpolation on a Chebyshev Grid is prefer rather than other interpolation methods?
4. In variational autoencoder experimental part, you should try more choices for hyperparameter N as it does in other experiment.
5. The experimental results in Table 2.1 show that IRDM is better in different dataset with different nodes, but it doesn't explain how to find a best hyperparameter N for different dataset.
6. Would you consider to re-organize the contents of chapter 3? As an independent chapter, do you think its technical contribution is enough? Perhaps you can extend the current version, or merge it into others.
7. Since it claims that the trained model more robust to large attacks, how do you define large attacks?
8. It may be better to provide more black-box attack setting details.
9. In addition to sketching, there are a group of low rank approximation methods for deep network compression. Perhaps you can discuss more in comparison with them.
10. Considering sketching is a main technique for the work in chapter 5, perhaps some other existing sketching methods can be discussed in more details.
11. Why do all the layers choose the same rank R_K for truncated_svd as in subsection (5.3.4), subsection (5.3.1), and algorithm 1? P_K in subsection (5.3.4).
12. In chapter 6, how to choose the parameter l? That is to say, how to find the number of pre-model layers?
13. Is it possible to analyze active subspace for all layers including pre-model ones, and get more compressed deep networks?
14. How does it perform by compare the proposed universal attack method with others counterparts?

15. To set a better bridge between the two parts on neural ODE and deep networks, is the sketching able to improve the performance of neural ODEs? is the active subspace analysis beneficial to neural ODEs?

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*