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: Maxim Rakhuba, Associate Professor, HSE University

I confirm the absence of any conflict of interest (Alternatively, Reviewer can formulate a possible conflict)	**Signature** **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

Evaluation of the thesis quality and overall structure of the dissertation

The Ph.D. thesis of Talgat Daulbaev consists of 6 chapters.

In Chapter 1, a general background is given. It includes the formulation of basic machine learning problems, neural ODEs, continuous normalizing flows, as well as reduced order modeling topics such as the discrete empirical interpolation method and the active subspaces method.

In Chapter 2, an interpolation-based method for fast and stable training of neural ODEs is introduced. One of the advantages of the proposed method is that the amount of the required memory is constant with respect to the number of ODE solver steps. As a result, the method shows consistent improvements in computing times over the standard adjoint method for neural ODEs on classification and density estimation tasks.

Chapter 3 contains an empirical study that tests the performance of different normalization techniques (batch, layer, spectral, and weight normalizations) for neural ODEs.

In Chapter 4, the author proposes a new approach for increasing the robustness of neural ODEs using Runge-Kutta methods (RKMs). The idea is to change the parameters of RKMs during training using different strategies. Two of the tested strategies (smoothing and switching) appear to be computationally efficient and show systematic improvements in the robustness of the resulting model in the conducted experiments.

Chapter 5 transfers ideas from model order reduction for differential equations to neural networks. In particular, the maximum volume algorithm is applied to the activations of neural networks in order to accelerate the inference. In some experiments on feed-forward neural networks on CIFAR-10 it is shown that the proposed approach is capable of speeding up computations up to 2.3 times without any accuracy drop.

In Chapter 6, the active subspace method is applied to neural networks in two ways. The first result is the acceleration and compression of neural networks by replacing a part of a neural network with a smaller one based on the active subspace method. The method is tested on CIFAR-10, where significant compression ratios (>20 times) and speedups (>7 times) are demonstrated with a < 3% accuracy drop. The second is a method for constructing universal adversarial attacks that involve computations of the dominant active subspace vector of a classifier. The proposed approach shows a ~20% higher attack ratio compared with a UAP baseline.

This thesis makes novel contributions, and it addresses important machine learning problems. The solutions are tested on commonly used benchmark datasets. Except for some issues with the consistency of formatting and misprints, the thesis is written well.

Regarding the drawbacks. Even though Chapter 3 is purely experimental, it does not contain many numerical tests. As a result, the analysis and conclusions are not particularly deep. In Chapter 5, there are no experiments on ResNets, which in a sense, are better motivated by neural ODEs than standard feed-forward architectures. Also, the results of Chapter 6 seem to be applicable to more modern architectures like transformers, but I understand that they were not very common at the time of publication.

The thesis contains multiple misprints, for example:

page 2: Algorith -> Algorithm

page 4: odel -> Model

page 16: see ??

page 39: black-ox -> black-box

page 52: formatting issue for the error

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

The content of this dissertation is definitely relevant to its actual topic. Chapters 2-4 cover neural ODEs, while chapters 5-6 are devoted to reduced order modeling. It is, however, worth noting that the title includes two topics (ODEs and reduced order modeling) that are not directly connected with each other. For example, Chapter 6 considers active subspaces that are not directly connected with differential equations.

The relevance of the methods used in the dissertation

In this thesis, methods from differential equations, numerical mathematics, and reduced-order modeling are applied to different deep learning problems. Most of the methods are state-of-the-art.

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

The scientific significance is in no doubt. The results are published in internationally recognized journals, including top publication venues. Most of the proposed methods are extensively tested on well-known benchmarks. Each chapter contains links to open-source implementations.

The relevance of the obtained results to applications

The results from Chapters 2-4 are applicable to a wide variety of problems where neural ODEs are used such as generative modeling. The acceleration methods from Chapters 5 and 6 are tested on computer vision problems, but they seem to be potentially useful in other application domains.

The quality of publications

All four publications by Talgat Daulbaev are of high quality. This includes one in a top conference in the field (NeurIPS), where Talgat is the first author, and one in a Q1 journal.

Conclusion

Overall, all remarks do not reduce the value of the work, and I recommend that the candidate should defend the thesis by means of a formal thesis 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*