Skoltech
Skolkovo Institute of Science and Technology

# Jury Member Report – Doctor of Philosophy thesis.

**Name of Candidate:** Valentin Khrulkov

**PhD Program:** Computational and Data Science and Engineering

**Title of Thesis:** Geometrical methods in machine learning and tensor analysis

**Supervisor:** Professor Ivan Oseledets, Skoltech

**Name of the Reviewer:**

| | |
|---|---|
| I confirm the absence of any conflict of interest<br><br>(Alternatively, Reviewer can formulate a possible conflict) | **Signature:**<br><br>27-09-2020<br>**Date: DD-MM-YYYY** |

*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

**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*

Technische Universität Berlin

TU Berlin | Straße des 17. Juni 136 | 10623 Berlin

**Fakultät II – Mathematik und Naturwissenschaften**
Institut für Mathematik

Angewandte Funktionalanalysis

Einstein-Professorin
**Prof. Dr. Gitta Kutyniok**

Sekr. MA 5-4
Raum MA 577
Straße des 17. Juni 136
10623 Berlin

Telefon +49 (0)30 314-25758
Telefax +49 (0)30 314-27364
kutyniok@math.tu-berlin.de

Skoltech Doctoral Program
Skolkovo Institute of Science and Technology
Bolshoy Boulevard 30, bld. 1
Moscow, Russia 121205

Berlin, 27. September 2020

Report on the PhD Thesis
"Geometrical Methods in Machine Learning and Tensor Analysis"
of Valentin Khrulkov

We currently witness the impressive success of deep neural networks in various real-world applications. Examples range from autonomous driving over game intelligence to the health care sector. A similarly strong impact of deep learning can be experienced in science, where previous model-based methods are now replaced by entirely data-driven approaches or hybrid-type algorithms. However, despite this outstanding success, a comprehensive theoretical foundation is still largely missing, but in great demand due to the sensitive nature of many of the applications. In fact, sometimes these techniques show a significant sensitivity to small perturbations, leading to dramatic failures such as in the so-called adversarial examples, thereby calling for a profound theory.

Neural networks were originally introduced in 1943 by McCulloch and Pitts as an approach to develop learning algorithms by mimicking the human brain. The key goal at that time was the introduction of a theory of artificial intelligence. However, the limited amount of data and the lack of high performance computers made the training of deep neural networks, i.e., networks with many layers, unfeasible. The true breakthrough of this approach came with the ability to train deep neural networks (instead of shallow ones) with large-scale training data sets.

Regarding deep learning as a statistical learning problem, three main directions of a theoretical underpinning can be identified according to the fact that the error of such a problem consists of three part: the approximation error related to the hypothesis class (here neural networks), the optimization error from the optimization procedure and the sampling error related to the availability of only sample values instead of the continuous distribution. Those directions are in deep learning language coined expressivity, learning, and generalization.

The thesis of Mr. Khrulkov addresses several key questions in those areas from a geometric perspective. In general, the thesis is a mathematically beautiful work, using deep mathematical techniques to make significant contributions to the theoretical foundation of deep learning, which at the same time are also practically relevant. The strength and importance of the results can also be seen from the fact that all related publications appear in top conferences or journals. Impressive is also the range of topics covered by this thesis, showing that Mr. Khrulkov not only masters one direction and methodological focus, say expressivity and approximation theory, but in fact an entire bouquet of directions. It is also most noteworthy that the mathematical (geometric) theories Mr. Khrulkov employs have often not been utilized in the area of the theory of deep learning before, thereby also opening the door for a novel set of tools for attacking theoretical problems of deep neural networks. Finally, the thesis is very well written with a well-designed structure, making the cumulative thesis very nice to read.

Let me now discuss the content of the thesis in more detail. Since it is a cumulative thesis with a nicely written overview, I will focus on two to my mind particularly interesting and important groups of results. The first concerns Mr. Khrulkov's results in the area of expressivity. As said, one key question is: Why are deep neural networks better than shallow ones? Approaching this question from the direction of expressivity, Mr. Khrulkov asks whether certain types of deep networks are equivalent to exponentially wider shallow networks, implying that those have exponentially more expressive power. Building on work by Cohen et al., he focusses on the class of recurrent neural networks with multiplicative nonlinearities. He then first shows that indeed the elements of this class have exponentially more expressive power than shallow networks. This is achieved by using tensor analysis, which provides a beautiful link between specific tensor decompositions and neural networks. In the case of recurrent neural networks, this is the CP decomposition. Thus, the key theorem implying the expressivity result states that a tensor in TT format has maximal canonical rank in the sense of the rank of its CP decomposition.
Since the condition of multiplicativity is quite restrictive, and in practise often other types of recurrent neural networks are used, Mr. Khrulkov then extends his results to the case of ReLU activation functions. This extension is highly non-trivial, since classical results from algebraic geometry are not anymore directly applicable. For these results, Mr. Khrulkov uses generalized tensor decompositions and, in contrast to the previous results, only compares the respective deep and shallow network on a finite grid; the later relying on the framework of grid tensors. Intriguingly, the results show a different picture than before in the sense that not all elements of the considered class have exponentially more expressive power than shallow networks; there exists an open set for which this is not true. This is somehow surprising, since one would expect that the customarily used ReLU networks are at least as good. Mr. Khrulkov's numerical experiments, which are carefully set up, analyze whether this impacts practise by estimating the rank of the equivalent shallow network. He observes that indeed this rank-problem occurs, but is diminished with increasing the rank of the generalized recurrent neural network. This is indeed a quite useful observation as well, and it would be interesting to discuss some first ideas how to make this theoretically precise.
In general, I regard these results as significant for a better understanding of recurrent neural networks. I would only like to point out that it would be advisable to also cite and interesting to compare with the work by Mhaskar et al., who approach this problem from the viewpoint of compositional functions.

The second quite intriguing contribution of Mr. Khrulkov I would like to highlight concerns the problem one faces in practise, namely the fact that the data manifold is typically unknown. This is usually attacked by using generative adversarial networks, which after training aim to generate new data points on the data manifold. However, one main problem is a suitable quality measure for the performance in the sense of how "good" the generated data points are. The mathematical theory of persistent homology provides a means to analyze data clouds and is the work horse in topological data analysis. In Mr. Khrulkov's approach to design a quality measure, Mr. Khrulkov brings these two worlds together, which I regard as quite innovative. Based on the key ingredients of persistent homology, namely simplicial complexes leading to persistent barcodes, he defines a novel quality measure for their topological similarity. One key idea is to estimate the probability distribution of the certainty of the number of holes of the data manifold on average.
He then tests this measure, which he coins Geometry Score, in carefully designed numerical experiments using two generative adversarial networks of which one is assumed to perform better. Besides the advantage that this new measure can be applied to any data set and not only imaging data, it is shown that it often also outperforms the previous measures. The only slight disadvantage – as Mr. Khrulkov correctly states himself – is that his measure does not evaluate the visual quality of the generated data points.

I regard Mr. Khrulkov's novel measure as quite innovative and opening a new avenue of evaluating generative adversarial networks. It would be nice to discuss in more detail how this could lead to, as it is stated, "further theoretical understanding of GANs". Moreover, it might be worth considering whether this measure could be combined with image quality measures such as the structural similarity index (SSIM).

Finally, throughout the thesis, I noticed some typos such as "nonlinearirities" on page 11; I would suggest to use a spell-checker. Moreover, the header is often too long and overlaps itself.

Concluding, this PhD thesis contains several significant contributions to the theoretical understanding of deep/machine learning from a geometrical perspective. Those are based on original, novel ideas and utilize deep and often non-standard mathematical theories and techniques in this area. Also, implications and limitations of the research are nicely discussed, and practical issues carefully taken into account and addressed.

Sincerely yours,