Skoltech
Skolkovo Institute of Science and Technology

Jury Member Report – Doctor of Philosophy thesis.

Name of Candidate: Marina Munkhoeva

PhD Program: Computational and Data Science and Engineering

Title of Thesis: Fast numerical linear algebra methods for machine learning

Supervisor: Professor Ivan Oseledets

Name of the Reviewer:

I confirm the absence of any conflict of interest (Alternatively, Reviewer can formulate a possible conflict)	**Signature:**  **Date: 20-March-2021**

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 broadly concerns the design of efficient numerical algorithms for machine learning problems dealing with high-dimensional data.

The thesis consists of 5 chapters, excluding introduction and conclusions.

Chapter 1 discusses the mathematical background.

Chapter 2 deals with numerical methods for kernel approximation and introduces a novel technique for quadrature-based random feature maps that is shown to generalise previous approaches such as random Fourier features.

Chapter 3 deals with the intrinsic geometry of the data and proposes a new distance between data manifolds based on stochastic Lanszos quadrature.

Chapter 4 deals with the approximation of spectral distances between graphs using matrix sketching algorithms, showing a linear complexity methods that scales to very large graphs.

Chapter 5 deals with graph embedding and proposes an "anytime" embedding algorithm with theoretical performance guarantees.

The novel scientific contributions are contained in Chapters 2-4. The candidate clearly shows knowledge of the current problems and state-of-the-art methods in the field of machine learning, and has an excellent command of modern numerical methods that she applies skillfully to address these problems. The latter, namely, the use of efficient numerical linear algebra techniques in ML, is leitmotif of her work. I am particularly pleased to see many geometric insights that I consider foundational for modern ML. Overall, the thesis is very well written, clearly organised, and reads very well.

The thesis is based on published (joint) first-author papers in top venues in the field, including a NeurIPS and ICLR paper. These are among the most competitive international conferences notoriously hard to get into.

I believe the work makes multiple contributions to the field, which fulfill the requirements for the PhD degree.

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*