

## 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:** Prof. Ivan Oseledets

**Name of the Reviewer:** Maxim Panov, Assistant Professor, CDISE, Skoltech

I confirm the absence of any conflict of interest.

**Signature:**



**Date: 08-02-2021**

*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 considered thesis targets a very important area of improving efficiency of machine learning methods via the application of fast numerical linear algebra methods. The author considers several important problems:

1. Kernel approximation via structured feature maps.
2. Measuring the distance between point clouds and graphs via spectral distances, and their efficient computation.
3. Construction of graph embeddings via efficient sketching algorithms.

All these topics are relevant for modern machine learning and data analysis. The author shows fluency in state-of-the-art numerical linear algebra methods and their applications to machine learning. Moreover, some of the methods require non-trivial modifications and lead to the algorithms, which are interesting for this area in general.

I also have some worries about the presentation of the results. First, consistently through the text it is very hard to distinguish between the statements that are well known in the literature and those that constitute the body of the main results in the thesis. It is especially worrying considering that all the papers are co-authored with other PhD students and thus a clear separation of research contributions is required. For example, Chapter 2 resembles the paper of Marina and Ermek Kapushev at NeurIPS conference, while Ermek is also about to defend his thesis. I highly recommend to clearly stating which results are obtained to Marina and which are going to be defended (in every chapter). The sentences as “In this thesis we propose ...” should significantly clarify the contributions of the work. In addition, I recommend to better link different parts of text with each other. For example, Section 2.4 appears suddenly without any explanation why is it needed. The same applies to Section 2.5 and some others. Speaking about Chapter 3, it would be great to have the formal problem statement as currently there is only informal description and then the author directly proceeds to heat kernels. In addition, I am not sure that I understand the sentence “data distance” in the title of the chapter. Probably, these two issues are related. If we continue to Chapter 5, then again problem statement is not entirely clear especially if we look on “Problem setting” section, which apparently does not contain problem statement.

In my opinion, thesis results are very relevant from the application perspective. Importantly, the author provides all the details on the algorithm construction and training procedures, which allows to directly applying them.

Finally, the results of the thesis research were published in proceedings of well-reputed conferences including two publications in CORE A\* conferences and one publication in highly prestigious ICLR conference which is not in CORE rating but has level equivalent to A\*. Thus, the quality of the publications well supports the overall good scientific quality of Marina’s thesis research.

While I have overall positive opinion about the research contents of the thesis I think that the text deserves some improvements (see some issues below). To sum up, I think that the issues found do not decrease the scientific quality of the thesis and Marina Munkhoeva deserves to be awarded with Skoltech PhD degree.

The list of issues (in the order of appearance in the text):

1. [list of publications] List of publications should contain full bibliographic references including pages.
2. [page 4] For the identity matrix condition  $I_{ij} = 0$  for  $i \neq j$  should be added.

3. [formula 1.1 and everywhere later] It is recommended to fully integrate formulas in the text, i.e. put commas or dots where needed (currently there is no consistency in this respect).
4. [Table 1.2] There is a mistake in the definition of normalized Laplacian.
5. [Table 2.2] For ROM and QF the complexity doesn't depend on D. How is it possible?
6. [Chapter 2] The pseudocode or listing of the resulting algorithm would be great to have. Currently, it is hard to decrypt it from the long text describing it.
7. [Definition 2, page 35] Somehow the limit for  $t \rightarrow 0$  results in function depending on t.
8. [Formula 3.16] It seems like the abbreviation slq first appears here and was not introduced before.
9. [Algorithm 1, page 43] Function IMDist probably was not meant for recursive application.
10. [Formula 5.1] Variable m seems to be not introduced

#### 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*