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 **I confirm there is no conflict of interest** (Alternatively, Reviewer can formulate a possible conflict)	**Signature:** Andrew Wilson **Date: 22-03-2021**

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 proposes methods for Gaussian processes, measuring distances between data manifolds (IMD), computing spectral graph distances, interpreting the non-linear VERSE embedding. Throughout, the methods are made scalable through pursuing ideas in numerical linear algebra. One of the key narratives in the thesis is that one does not need to (significantly) sacrifice performance for scalability if we can develop algebraic approaches to scaling. Applications include standard regression and classification for Gaussian processes, and several unsupervised tasks for IMD, such as evaluating GAN image samples, and tracking the evolution of image manifolds.

The thesis was in general a pleasure to read. It is nicely written, and I wholeheartedly endorse the narrative that we need not trade-off the fidelity of our model representation for scalability and that numerical linear algebra provides a foundation for achieving this ambition. I do not believe there are any serious issues with this thesis. The comments below are largely "future directions" rather than necessary corrections.

In many cases, the assumptions already made by existing methods — such as popular choices for kernels for Gaussian processes — can be immediately exploited for scalable inference without introducing additional approximations [e.g., 1,2]. Alternatively, one can develop Krylov subspace methods in conjunction with GPU parallelization for scalable *exact* approaches [e.g., 3]. In other words, often there can be no trade-off at all. In many of the methods proposed in this thesis, there is still a trade-off or an additional approximation, which can be unnecessary. It would be useful to see comparisons between the proposed quadrature approach and for example the methods in GPyTorch [3].

Indeed, for many of the proposed approaches, the practical utility — and extent of practical applicability — is unclear. For instance, in what particular application settings would we prefer the quadrature GP approach? What are the limitations with dimensionality? With various kernel types? As a *global* approximation, compared to local approximations such as FITC, it seems the quadrature approach could be promising in tasks such as Bayesian optimization, where we do not know a priori where we need our approximation to be accurate. But there were no such downstream tasks presented. Similarly, the data manifold methods are demonstrated in more of a "proof of concept" fashion on very qualitative tasks, like GAN image quality evaluation. Could this method not be useful for a downstream task like neural architecture search? It would be nice to see where these methods could actually make a practical difference.

Krylov methods such as stochastic Lanczos quadrature surface regularly throughout the thesis. A key advantage of these methods is that they are easy to parallelize, and can thus be greatly accelerated by modern hardware such as GPUs. Ideally, it would have been good to have done studies on how GPU computing can make the proposed methods particularly compelling. It will be important to at least add a *discussion* of how the proposed methods can be parallelized (in some cases relative to alternative approximation procedures that are more sequential) and how parallelization of this kind is particularly relevant given advances in hardware design.

[1] A.G. Wilson, E. Gilboa, A. Nehorai, and J.P. Cunningham. Fast kernel learning for multidimensional pattern extrapolation. Advances in Neural Information Processing Systems, 2014.
[2] A.G. Wilson and H. Nickisch. Kernel Interpolation for Scalable Structured Gaussian processes (KISS-GP). International Conference on Machine Learning. 2015.
[3] J. Gardner, G. Pleiss, D. Bindel, K. Weinberger, A.G. Wilson. GPyTorch: Blackbox Matrix- Matrix Gaussian Process Inference with GPU Acceleration. Advances in Neural Information Processing Systems, 2018.

Provisional Recommendation
I recommend that the candidate should defend the thesis by means of a formal thesis defense. *<— My recommendation*
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