Skoltech
Skolkovo Institute of Science and Technology

Jury Member Report – Doctor of Philosophy thesis.

Name of Candidate: Vladimir Fanaskov

PhD Program: Mathematics and Mechanics

Title of Thesis: Statistical inference and machine learning in numerical linear algebra

Supervisor: Associate Professor Aslan Kasimov

Name of the Reviewer: Nikolai Brilliantov

I confirm the absence of any conflict of interest ~~(Alternatively, Reviewer can formulate a possible conflict)~~	**Date:14.08.2022**

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 present thesis is devoted to the numerical solution of linear systems, Ax=b (x-is the unknown vector, A is the matrix of coefficients and b-known right-hand side), which is an important problem in many domains of science. These include optimization, control, probability, and scientific computing, just to mention a few. The author focuses on a particular type of linear systems with a matrix A that contains multiple zeros (or has a structure associated with convolution) to make the multiplication of matrix by vector ("matvec") computationally cheap. This kind of matrices often appear in integral, integro-differential, and partial differential equations when finite difference or finite element discretizations are utilized. In particular, most of the examples addressed by the applicant refer to the mechanics of continuum media. More specifically, he considers heat equation, Helmholtz equation, Poisson equation, biharmonic equation, convection-diffusion equation, equations with sharp boundary layers, etc. All that makes obtained results practically relevant and applicable to simulations of real engineering problems.

Being important, the problem of numerically approximating the solution of linear systems has been extensively studied. The main novelty of the present thesis is in the use of machine learning and statistical techniques for the construction of numerical linear systems. Despite of its efficiency, such techniques have received much less attention in the past as compared with the more traditional ones. The main line of the developed approach is as follows. Since the computational algorithm "matvec" is cheap, the iterative techniques are competitive with the direct solvers. Still, the construction of iterative solvers can be complicated, hence the candidate suggests utilizing the Bayesian statistics and machine learning. More precisely, this has been done by optimization, learning, as well as using the analogy the statistical approach and the solution to linear systems (mostly in the context of multivariate normal distribution).

Chosen approaches logically divide the thesis into two main parts. In the first part, statistical techniques prevail whereas the second part is based on the machine learning approaches.

In the first part, the candidate extensively exploits the relation between the multivariate normal models and linear problems: In the second chapter, he uses the fact, that for certain normal statistical models (a particular Gaussian Markov random field) marginal distributions contain information about the solution of a linear system. The resulting algorithm is then further generalized to non-symmetric systems for which the initial statistical model is not well defined. Here, the analogy with the Gaussian elimination is used and statistical interpretation is abandoned. The third chapter is also built on a multivariate normal model approach, for which the mean of the conditional distribution is related to the projection methods (the so-called "Krylov subspace methods") and the variance measures the uncertainty of the solution. The goal is to construct a prior in such a way that the posterior generates an exact solution with high probability. Here a novel statistical model is proposed that outperforms the existing ones. That is, it provides better uncertainty quantification for an error of the approximate solution generated by the projection method. The technique applies to a broad family of projection methods including conjugate gradient, GMRES, and their preconditioned variants. In the fourth chapter the candidate explains yet another original uncertainty quantification technique for the solution of linear systems. In this case the normal model is not involved in the most general settings. Nevertheless, it appears later, when variational inference is used to make posterior computationally tractable. The results of this chapter were also used to accelerate the relaxation solver with additional projection steps on low dimensional subspaces collinear to the error vector. The subspaces are constructed during the uncertainty quantification step.

The second part is of the thesis is of particular interest, since it presents methods allowing the construction of linear solvers and preconditioners in an automatic ("black-box") or semi-automatic mode. In the three main chapters of the second part the candidate utilizes a standard machine learning pipeline. Namely, an appropriate loss function is introduced, architecture is defined and free parameters (e.g., weights of the neural networks) are optimized using the method from the stochastic gradient descent family. The exploited architecture in chapter 6, is essentially a generalization of the famous BPX preconditioners and the loss function is a stochastic trace approximation to the spectral radius of a preconditioned linear system.

In chapter 7, the architecture used is an in-house operator-free multigrid solver, and the loss function is the same approximation to spectral radius as before, however the matrix under consideration is the error propagation matrix. It is also shown here that fixed architectures can be hardly generalized when the grid is substantially refined. In other words, the performance of a trained model can become arbitrarily poor with an increasing number of grid points. Candidate has proposed to resolve this problem using a serialization of layers. The performance of the proposed architectures compares favourably with the U-Net solvers and previously developed neural-multigrid architectures. In chapter 8 stays the technique of reinforcement learning is exploited for the online optimization of iterative methods (relaxation methods and multigrid). Here the free parameters are relaxation parameters (e.g., weights in successive over-relaxation). The results of the chapter show that the reinforcement learning can be successfully used in linear algebra to construct the adaptive iterative solvers that simultaneously solve a given set of linear problem and improve the convergence speed. Here the byproduct such as the norm of the residuals to guide self-optimizations is exploited.

In chapters 9-11, which comprise the Appendix to the thesis some technical detail, examples of the partial differential equations (PDE) and the proof of theorems is provided. I find this part of the thesis is also very useful.

All in all, theoretical results or/and conjectures presented in the thesis are substantiated either by direct proofs or by numerical experiments. Proposed algorithms are benchmarked against a large number of state-of-the-art linear solvers (geometric multigrid, polynomial, and color relaxation techniques, projection techniques, preconditioned iterations, and decompositions) and uncertainty quantification techniques. As explained above, linear systems used by the author originate from the discretization of PDEs which are benchmarks of choice for iterative methods. Besides that, linear solvers are a major bottleneck in the numerical solution of PDE which makes thesis results useful for applications.

The results of the thesis have been published in three prestigious international journals, among which is SISC -- one of the best journals in scientific computing. The results have been also reported at two leading international conferences. Most of the articles are single-authored, except for the article on BPX, written together with one co-author. The thesis contains many interesting and novel scientific results, and seems to be acceptable as it is, after minor modification indicated below.

The comments mainly refer to the exposition of the material (not in the order of importance):

1)I suggest expanding the introduction, adding a more detailed explanation, how different approaches used in the thesis are related to each other and to the solution of linear problems in general. Moreover, the author should invest for efforts to make the introduction chapter more readable for non-experts, probably adding more simple explanatory examples.

2)I believe that the reference to the Boltzmann distribution in Eq. (2.6) is misleading (there are no important parameter, temperature there). Hence it should be either explained in more detail or removed.

3)It would be worth to add a special section where the discussion about the limitations of the applied methods should be given. I admit that such a discussion is distributed piecewise through the thesis, however I suggest accumulating them in one place, possibly in the conclusion.

4) Please make a careful proofreading of the thesis text. There exist misprints and errors. For instance, in Eq. (10.11) the factor (1/h^2) is missed, and in Eq. (10.17) there is a misprint “y2”. Besides, the meaning of Eq. (10.17) is questionable: for both cases, i=1,2 the same result is obtained.

Provisional Recommendation
<u>I recommend that the candidate should defend the thesis by means of a formal thesis defense</u>
~~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~~