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:

I confirm the absence of any conflict of interest (Alternatively, Reviewer can formulate a possible conflict)	**01-09-2022** **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

The thesis consists of two parts, 8 chapters, presents the various research topics: probabilistic methods for solving sparse linear system, optimization methods to find good BPX preconditioners, Deep NN architectures which are suitable for construction of linear solvers. The thesis also provides appendices which contain proofs of theorems. Plenty numerical examples are presented to demonstrate performances of the proposed algorithms and methods.

Detailed comments: The author can consider the following comments to improve the clarity of the thesis.

- Introduction section presents the goal of the research, e.g., study relations between statistical inference, machine learning and iterative algorithms, study probabilistic numerical algorithms for solving sparse linear systems. The goal and contribution of the thesis are clear. However, the motivation of the research and formulation of the studied problems are missing. Why are the studied problems important, e.g., probabilistic algorithms and generalization of BPX conditioners using machine learning methods, applying Bayesian analysis to solving linear systems? What are the benefits of the proposed algorithms, e.g., the probabilistic Richardson algorithm?

- Appendices 9 and 10 should be presented in the Introduction or Preliminary section or a section for formulated problems. This will help the readers to understand an overview of the challenging problems considered in the thesis.

Important concepts should be presented in the preliminary section, e.g., ex- isting probabilistic algorithms, well and poorly calibrated system, ”instationary” Richardson algorithm, multigrid method based on Deep NNs.

- Contribution ”Neural architecture is equivalent to the multigrid solver” and Section 7

”Neural architecture” is not a research topic, but neural architecture search or neural architecture design for a specific task.

- Inconsistent notation for matrices, vectors, normal distribution

+ Boldface for matrices on page 16, but italic font on pages 14, 23, (2.32) on page 33, Algorithm 2, error propagation matrix M on page 37

+ Normal distribution on page 79 $N(\mu,\Sigma)$ and page 13 $N(\mu,\Sigma)$

-Parallel version (implementation) of the proposed GabP algorithm, pages 28. Do the two implementations converge to the same solution? Is the parallel version faster than Algorithm 1?

-Frontpage ”Moscow 2021” -> ”2022”

- Confusing notation: $|\tilde{R}|$ in Theorem 2.3.2 can be understood as determinant of the matrix $\tilde{R}$.

- Theorem 28. Does δ_{ij} denote the Kronecker delta? The notation is not defined in Chapter 2.

- Page 38, results shown in Figure 2.4.
Why is GaBP robust? GaBP does not converge faster than Gauss-Seidel. GaBp(k) with different sweeps k = 1, 2, 3 do not have the same or comparable

complexity to update x^m per iteration. Hence it is nonsense to compare converge of GaBP(k) with different k and Gauss-Seidel.

There is no sweep parameter k in Algorithms 1 and 2.
x^n, $x^{(n)}$ and x^m: Why don’t use one notation?
- Table 2.3: What are the numbers shown in each cell in Table 2.3?
- Summary on page 43
”Red-black has better convergence rate than other solvers”
Again, since different solvers have different complexity per iteration, the

number of iterations is not an appropriate measure for the convergence rate. For example GaBP(3) with k = 3 sweeps should demand lower number of ”outer” iterations than GaBP(2), but it does not mean that GaBP(3) converges faster than GaBP(2).

- Page 48:
What is the problem of probabilistic reconstruction?
”Lanczost algorithm” shall be ”Lanczos” algorithm.
- Page 50: $W^T AW$
A should be A.
-Page 56
Why is the factor s common?

-Page 58, definition of S-statistic in Lemma 3.3.5
The first term $(x - \tilde{x})$ misses a transpose operator.
- Page 65
G = I should be denoted as two matrices.
- Figure 4.3 on page 77 should be moved to where it is first mentioned.
- Page 78
Check notation of v_0 and matrices A, V
- Page 87
"instationary": should be "nonstationary""
"N = I" should be denoted as matrices
- Page 88
What does the symbol ⊗ stand for?
Check notation for vector μ and matrix Σ in (4.48) and (4.49)
- Page 89
Move Algorithm 7 to page 93 or where this algorithm is first mentioned.
- Part 1 misses a section for conclusion. What is the benefit of the probabilis-

tic algorithm over deterministic algorithms? When can the proposed algorithm be applied?

- Part 2, page 101

"The rest part ... provides more examples on how machine learning can be used in numeral linear algebra""

What examples? please be more specific.
- Page 103
What does B_{BPX} in (6.5) improve the pre-conditioner over that in (6.4)?
- Page 105
Check notation for the matrices R and A in section 6.1.2: here R is an easy

invertible approximation to A
-Page 109, (6.14)
What are the differences between matrices S_l, for l = 1,2,...? The same

question is for matrices I_l. - Page 111

Where is Figure 6.3a? There is a table with caption Figure 6.3 on page 114 but no figure.

- Numerical examples and the whole Chapter 6, it is not clear how Machine learning methods are applied to find the optimal preconditioner. Does the au- thor mean that minimization of spectral conditional number uses ML methods?

Figures 6.3 6.4 and 6.5 are tables without caption. Check the captions or provide figures instead of tables.

- Chapter 7

Similar to Figures in Chapter 6, Figures 7.3, 7.4, 7.5 and 7.6 are tables. Move theses tables to the appropriate places. Captions of theses tables or figures should be self-contained

- Chapter 8

The over-relaxation parameter w is not defined yet on page 132. What is the role of w in the SOR? Section 8.1 should formulate the problem with SOR before introducing the task of finding accurate approximation to w.

Why does the agent go with her on page 183? Page 134, what are R1, ..., RN in (8.2)? ”MDS” ?
- Page 135, Algorithm 11

What is outcome of Algorithm 11?

The running index i changes inside the for loop at Line 6 or Line 8. This affects the output at line 11.

- Page 136

Provide applications where we need to solve the same linear system more than once.

- Page 139, Algorithms 12 and 13

M(θ) denotes the error propagation matrix, but in Algorithm 12, the same notation represents the iterative method.

Line 4 and Line 11, ‖x−y| should be ‖x−y‖
- Page 140, Figure 8.1
Dashed and solid lines in Figure 8.1 and Figure 8.2 are not distinguished. - Other mistakes, grammatical errors ”Both of this routes” page 14 ”Algorithm 4 simply perform an additional run”
”are already present” on page 137

”This parameters” (page 137)
”we drawn 50 matrices” (page 138)
(drew)
page 145
”we on a more serious examples we perform a series of ...”

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*