Skolkovo Institute of Science and Technology

Jury Member Report – Doctor of Philosophy thesis.

Name of Candidate: Georgii Novikov

PhD Program: Computational and Data Science and Engineering

Title of Thesis: Tensor-train decomposition for efficient data representation and learning

Supervisor: Professor Ivan Oseledets
Co-supervisor: Assistant Professor Maxim Panov

Name of the Reviewer: Petr Tichavsky, Research Professor

I confirm the absence of any conflict of interest	**Date: 15-11-2024**

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 consists of five parts.

Chapter 1 contains a necessary introduction to tensor train (TT) modeling. It covers operations on tensors in TT form, conversion of the full tensor into TT via singular value decomposition (SVD), TT rounding (re-compression), and a few optimization methods of a tensor in the TT format: alternating least squares and the Riemannian gradient computation. These concepts will be used in further developments. Section 1.4 contains a concept of the tensor-train matrix. It is used later in Section 4. I remark in advance that the tensor train matrix is not much different from the TT model. I will explain this remark later in more detail.

In general, Chapter 1 is well-written and easy to read. Maybe except for section 1.2.4 which is a bit incomplete and therefore confusing. It contains definitions of left-orthogonal core, right-orthogonal core, it claims the existence of mu-orthogonal decomposition, but it does not provide details on how a general TT can be converted into another one with orthogonal cores and how mu-orthogonal decomposition can be found. Algorithm 1.2 promises cores $G^{(i)}$ of the TT approximation of the tensor with the given rank as the output, but it returns some cores with different notations, $U^{(i)}$. It isn't clear at least. Chapter 1 is not essential for the Thesis, the novel parts are mainly in Chapters 2-5.

Chapter 2 presents a way to model multivariate probability density using TT. To the best of my knowledge, this interesting idea is novel. I'd be, however, afraid of possible negativity of the probability density. Therefore I would suggest putting more focus on the squared TTDE, maybe even starting with this. In other words, I would suggest to follow quantum mechanics and model a wave function as a TT. The wave function could be possibly complex-valued. The probability density would be the squared module of the wave function. The advantage is that the normalization can be done easily. It is easy to compute the Frobenius norm of a tensor in TT format.

I would like to see the SqrTTDE operations mentioned on page 56 in more detail, in equations, not only in words. Section 2.2.3 on sampling is fine. Section 2.3. is done for TTDE, not for SqrTTDE, correct? My question is if it would be feasible to do it for SqrTTDE as well. But still, it is a nice piece of work.

Chapter 3 presents another interesting application of TT which is the modeling of fully-connected layers in neural networks. The model with TT matrix representation is very appealing and nice. Only I can't agree with the author that it has to be learned from scratch. (The first sentence of section 3.4. on page 91). We can model the matrix as an ordinary TT, with only a little pre-processing. If a learned fully connected layer is given, we can reshape it to a tensor as in section 3.2.1 on page 81. Then, we have to permute the tensor coordinates in the order (1,k+1,2,k+2,....,k,2k), reshape it again to order-k tensor,and then we can use the TT-SVD algorithm to identify the core tensor. However, this comment is of minor importance only. The Chapter contains a very interesting method of replacing the fully-connected layers. Experiments show its usefulness. If, however, we decompose a learned fully connected layer, we would get important information on what bond dimensions are necessary.

Chapter 4 presents an interesting material as well. It is about approximating nonlinear functions by piecewise constant functions. This part is not well developed. When Dynamic programming is applied, eq. (4.120), some reference should be provided, and more explanations are needed, otherwise, it is not clear what the author is doing.

Chapter 5 presents an application of TT decomposition in approximate neighbor search in large vector databases. At least it looks like that, at the first glance. After defining the TT approximation of the matrix Y on page 125, it is written that direct TT compression of tensorized matrix Y_{TT} gives unacceptable results. The remainder of the section is not quite clear to me, namely the probabilistic interpretation. The probability distribution p_Y is unknown. OK. I understand Y_{TT} but I can't understand how Y_{TT} induces a discrete finite distribution $p_{Y_{TT}}$, how this is defined. Moreover, Y_{TT} can be a poor approximation of Y, correct? Next, what is the standard SGD method? (What is the acronym SGD ?) I understand the Sliced Wasserstein Loss and the Nearest-Neighbor Distance.

The method proposed by the author is quite complex with several heuristics involved. See Algorithm 5.10 on page 139. It is not fortunate to have that long comments included in the table. They should be placed outside. It is good that it outperforms the (G)NO-IMI clustering method. It is not clear to me if the improvement is thanks to the TT model used or thanks to the Sliced Wasserstein Loss (the latter loss is not mentioned in [83]). Could the author comment on this?

Typos: memroy (bottom of page 136). In general, typos are very rare in the Thesis.

Overall evaluation:

The Thesis contains some not quite clear parts but still, it includes very valuable material (namely in Chapters 1-3) for which the author deserves to get his PhD.

- *Brief evaluation of the thesis quality and overall structure of the dissertation.*
  The thesis has a varying quality. The first three chapters are of very good quality. The chapters 4 and 5 are less clear, probably because they were not so deeply reviewed in publications before.
- *The relevance of the topic of dissertation work to its actual content*
  Chapters 1-3 are highly relevant for the topic of dissertation, Chapters 4 and 5 are less relevant. In my opinion, Chapters 1-3 would be sufficient.
- *The relevance of the methods used in the dissertation*
  The methods are highly relevant.
- *The scientific significance of the results obtained and their compliance with the international level and current state of the art*
  The scientific significance is high. I appreciate namely the modeling of probability density functions, which can have many applications, and compression of the fully connected layers in neural networks.
- *The relevance of the obtained results to applications (if applicable)*
  In my opinion, the results are applicable in practice. For example, in neural networks.
- *The quality of publications*
  Excellent.
- *The summary of issues to be addressed before/during the thesis defense*
  The author may try to address my comments in my review.

Provisional Recommendation

☒ *I recommend that the candidate should defend the thesis by means of a formal thesis defense*

Prague, Czech Republic, November 15, 2024. Petr Tichavsky