Supervisor's Report

Title: **Mixed Precision in Uncertainty Quantification Methods**
Author: **Josef Martínek**

Summary of the work's content

This thesis revolves around the computational solution of partial differential equations (PDEs) arising from real-world physical problems. In practice, the physical problems and thus the resulting PDEs often naturally contain uncertainties in the input data. It is thus critical to study how and to what extent uncertainties in the input affect the resulting solution. Answering these questions is the goal of the field of uncertainty quantification.

Monte Carlo methods are commonly-used approaches in uncertainty quantification for estimating a given quantity of interest. The so-called Multi-level Monte Carlo (MLMC) method is a Monte Carlo-based approach that makes use of a hierarchy of discretization levels; sampling on this hierarchy of levels serves to reduce the overall computational cost of computing the estimator, where the number of samples required on each level may be chosen adaptively.

Ultimately, these methods involve solving systems of linear algebraic equations associated with the various levels, which can be a computational bottleneck. The state-of-the-art in almost all applications, including uncertainty quantification, has thus far been to use the standard IEEE double precision (64 bits) for the solution of these linear systems. However, the increasing commercial availability of low precision hardware in modern high-performance computers has motivated a surge of research into where and how low precision can be used in computations without affecting accuracy. The use of lower precision has the potential to significantly increase performance, since the computer is moving and computing with many fewer bits. For example, on the latest NVIDIA H100 GPU, which contains the range from double to quarter (8 bit) precision, single precision (32 bits) is around 2× faster than double precision, and half precision (16 bits) is around 4× faster than double precision. Thus, if low precision could be used within MLMC methods without affecting the accuracy, this could result in significant practical speedups.

In his thesis, Josef Martínek has performed the first rigorous analysis of the use of mixed precision within MLMC methods for uncertainty quantification. After giving a thorough background on PDEs, Monte Carlo methods for uncertainty quantification, and finite precision numerical linear algebra, Josef Martínek has derived a complete convergence analysis of Monte Carlo methods in finite precision arithmetic via error estimates. This analysis is used to give insight into when and where low precision arithmetic can be exploited to improve performance.

The thesis then goes on to develop a mixed precision MLMC method, called MPMLMC, which aims to use lower precisions selectively in order to minimize the computational cost while still providing results as accurately and reliably as the uniform precision approach. The intuition is that the coarser the level of discretization, the lower the precision that can safely be used without detrimental effect. This intuition is confirmed by Josef Martínek's rigorous analysis. The utility of the theory is illustrated through a set of numerical experiments on model problems, which show that this approach can result in a 4-8× theoretical speedup.

Evaluation

- **Thesis topic:** The thesis topic is adequate for the Master's thesis level and the student has fulfilled the outlined requirements.
- **Author contribution:** Josef Martínek's results and analysis give, to my knowledge, the first rigorous justification that mixed precision can be safely used within uncertainty quantification methods to improve runtime. This is a major step forward and opens the door to the future

investigation into and theoretical analysis of the use of low and mixed precision within other approaches for uncertainty quantification and, in general, in the numerical solution of PDEs. I have no doubt that his work will inspire and lead to further breakthroughs in the use of mixed precision within uncertainty quantification methods.

- **Mathematics level:** In my opinion, the level of mathematics goes beyond that typically required for a Master's thesis. The work represents the successful combination of various disciplines, touching on aspects of mathematical modeling, uncertainty quantification, high-performance computing, and numerical linear algebra. I consider the work to be highly successful and particularly novel for a Master's thesis, and I suspect that it will lead to a quality publication in a high-impact journal.
- **Literature and sources:** The sources used are correctly cited, and include a good mix of text-books, classical articles, and more recent articles.
- **English and writing:** The quality of English language and grammar is acceptable.

Conclusion

I recommend accepting the presented thesis as a Master's Thesis.

in Prague, 30.05.2023

Dr. Erin Claire Carson, Ph.D.