Zeroth-order optimization for LLM Fine-Tuning

Veprikov Andrey
MIPT, Moscow
ISP RAS, Moscow
veprikov.as@phystech.edu

Grigoriy Evseev
MIPT, Moscow
evseev.gv@phystech.edu

Egor Petrov
MIPT, Moscow
petrov.egor.d@phystech.edu

Aleksandr Beznosikov
MIPT, Moscow
Innopolis University, Innopolis
anbeznosikov@gmail.com

Abstract

In the field of natural language processing, the standard approach is to pre-train large language models (LLMs) using first-order optimization techniques such as SGD and Adam. However, as the size of LLMs increases, the significant memory overhead associated with back-propagation to compute gradients becomes a serious problem due to insufficient memory for training. For this reason, more and more zeroth-order optimization (ZO) methods are being developed, which only require forward pass of the model to compute gradients. In this paper, we present a new ZO approach for LLM pre-training, and compare it with existing methods such as ZO-SGD and ZO-Adam.

Keywords: Zeroth-Order Optimization, Large Language Models (LLMs), Fine-Tuning, Machine Learning.
Highlights:

1. Novel Zeroth-Order Optimization Method
   A new zeroth-order optimization method developed for fine-tuning large language models (LLMs) is proposed.
2. Comparative Analysis
   A detailed comparison with existing zeroth-order methods such as ZO-SGD and ZO-Adam is made.
3. Memory and Computational Efficiency
   Experimentally verified X% reduction in memory utilization compared to traditional methods such as SGD and Adam.

1 Introduction

2 Main Results

3 Conclusion

4 Experiments