

## Abstract

This paper analyzes a spectral momentum optimizer for transformers that achieved worse performance (loss=9.85) compared to standard baselines like AdamW (loss=4.93). We identify key challenges in applying spectral methods to transformer optimization.

## 1 Introduction

Recent work has shown...

## 2 Method

Key components:

- Block-sparse attention
- Mixed precision training
- Newton-Schulz orthogonalization

## 3 Results

Method	Loss
SpectralMuon	9.85
AdamW	4.93
Muon	3.54

## 4 Conclusion

Future work should investigate...