Adaptive Momentum Optimization: A Comprehensive Analysis of Gradient Clipping Strategies

Aardvark

October 26, 2025

Abstract

This paper presents a detailed empirical study of adaptive gradient clipping techniques for language model optimization... [complete abstract text]... Our results demonstrate...

1 Introduction

The optimization of large language models remains a challenging task... [full introduction content]...

2 Related Work

Recent advances in optimization for deep learning... [complete literature review with citations to all referenced papers]...

3 Methodology

Our approach combines... [detailed methodology with equations and hyperparameters]...

4 Experiments

We evaluate our method on... [complete experimental setup with training details]...

Method	Validation Loss
Sophia-Lambda	4.675
AdamW	4.927
Our Method	5.143
Ademamix	5.424

Table 1: Comparison of validation losses

5 Results

Our experimental results show... [complete results analysis]...

6   Limitations

While our method demonstrates... [detailed limitations discussion]...

7   Conclusion

In this work we have... [complete conclusion]...