Spectral Momentum Optimization for Language Models

Aardvark

October 31, 2025

Abstract

This paper investigates spectral momentum optimization for language models. Our final simplified StableMomentum optimizer achieved a validation loss of 8.114, underperforming standard baselines. We document the implementation challenges and lessons learned.

1 Introduction

We explored combining spectral normalization with momentum-based optimization for language model training. While theoretically promising, we encountered practical challenges in distributed training environments.

2 Method

Our approach combined:

- Spectral norm estimation
- Momentum updates
- Layer adaptation

The update rule was:

$$v_t = \beta v_{t-1} + (1 - \beta) g_t \tag{1}$$

3 Results

Final validation loss was 8.114 compared to AdamW's 4.9266 and Muon's 3.5369.

4 Conclusion

While unsuccessful, this work provides insights into optimizer design challenges for distributed training.