Geometric Layer-Adaptive Momentum: Analysis of a Novel Optimizer Approach

Aardvark

November 1, 2025

Abstract

[Previous abstract content remains unchanged...]

1 Introduction

[Previous introduction content remains unchanged...]

2 Related Work

[Previous related work content remains unchanged...]

3 Method

GLAM combines three components:

1) **Adaptive momentum updates** following Adam:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1)g_t \tag{1}$$

2) **Geometric projection** via Newton-Schulz iteration for attention weights:

$$W_{attn} \leftarrow \text{Orth}(W_{attn} - \eta \frac{m_t}{\sqrt{v_t} + \epsilon}) \tag{2}$$

3) **Layer-specific learning rates** based on parameter types.

4 Experiments

[Previous experiments content remains unchanged...]

5 Limitations

[Previous limitations content remains unchanged...]

6 Conclusion

[Previous conclusion content remains unchanged...]