Why Geometric Transformations Underperform in Transformer Feedforward Networks

Aardvark

October 30, 2025

Abstract

This paper investigates geometric transformations in transformer feedforward networks. Our experiments show these approaches underperform standard designs.

1 Results

Our geometric feedforward variants achieved:

- Full-rank: 5.657 validation loss (40.1GB memory)

- Low-rank: 5.123 validation loss (32.5GB memory)

Compared to SwiGLU baseline: 4.927 (31.5GB memory).

2 Conclusion

Geometric transformations underperform standard feedforward designs in practice.