

Preliminary Hyperparameter Sweep (2 GPU, H100)

Branch: sweep

Date: December 6, 2024

Hardware: 2× NVIDIA H100 80GB HBM3

Objective: Identify optimal momentum gradient for layer-wise momentum decay (LMD) by testing three different slopes against baseline.

Experimental Design

This preliminary sweep tested whether varying momentum across transformer layers could improve training efficiency. The hypothesis was that deep layers (learning foundational features) should use high momentum for stability, while shallow layers (learning task-specific features) should use low momentum for adaptability.

Configurations tested:

- **Baseline:** Uniform momentum $\beta=0.95$ across all 12 layers
- **Conservative:** Linear momentum ramp from $\beta=0.96$ (layer 0) $\rightarrow \beta=0.94$ (layer 11)
- **Moderate:** Linear momentum ramp from $\beta=0.98$ (layer 0) $\rightarrow \beta=0.90$ (layer 11)
- **Aggressive:** Linear momentum ramp from $\beta=0.99$ (layer 0) $\rightarrow \beta=0.85$ (layer 11)
- **Inverse (control):** Linear momentum ramp from $\beta=0.90$ (layer 0) $\rightarrow \beta=0.98$ (layer 11) - tests if directionality matters

Training details:

- Model: NanoGPT (124M parameters, 12 layers, 6 heads, 768 embedding dim)
- Optimizer: Muon (MomentUm Orthogonalized by Newton-Schulz) for transformer blocks, AdamW for output head
- Dataset: FineWeb-10B
- Iterations: 5,100 steps
- Global batch size: 128 (64 per GPU \times 2 GPUs)
- Learning rate: 0.0036 (base), 0.00036 (Muon blocks)
- Validation frequency: Every 125 steps

Log Files

Configuration	Seeds	Log Files
Baseline	983	baseline_seed983_d752d086.txt
Conservative	41512, 54740, 57754	conservative_seed41512_86bb0002.txt conservative_seed54740_54def01b.txt

		conservative_seed57754_5cc438d4.txt
Moderate	42, 112, 25392, 86050	moderate_seed42_86bb0002.txt moderate_seed112_86bb0002.txt moderate_seed25392_403154ed.txt moderate_seed86050_9bf5e2ac.txt
Aggressive	32514, 79280, 91359	aggressive_seed32514_c9f8e8d0.txt aggressive_seed79280_78201597.txt aggressive_seed91359_d92e4c3b.txt
Inverse	12564, 71486, 82214	inverse_seed12564_a49d0ae0.txt inverse_seed71486_53310e75.txt inverse_seed82214_8779476d.txt