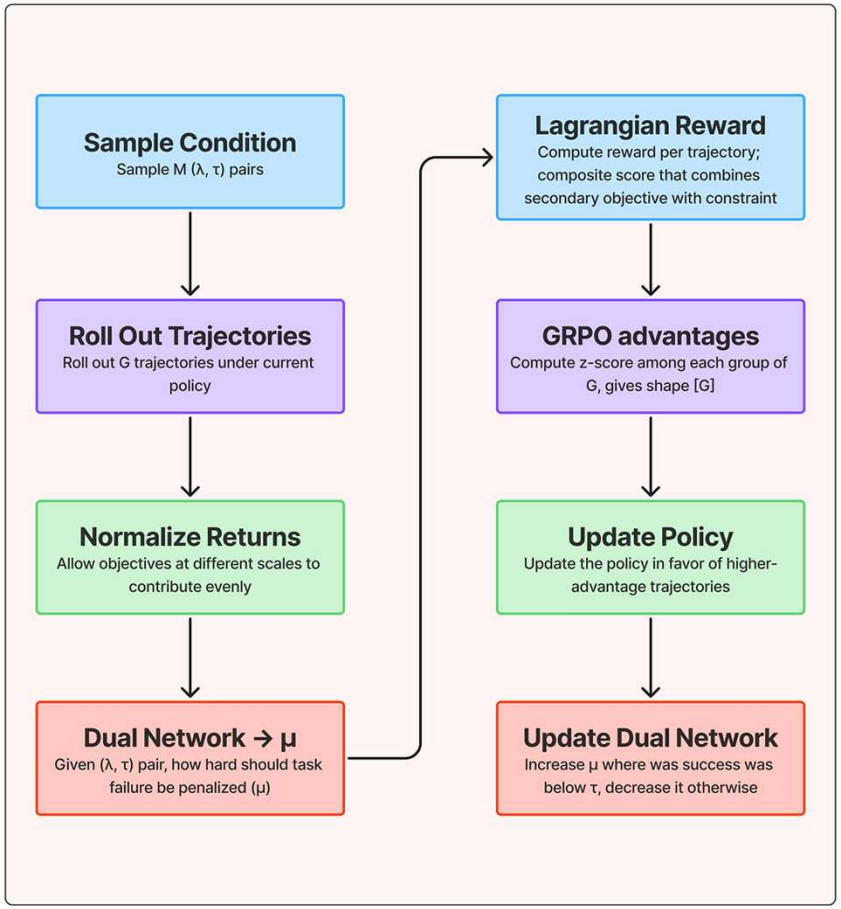
WEEK 5 UPDATE

CODY HILL • 6/22/26

Problem Overview

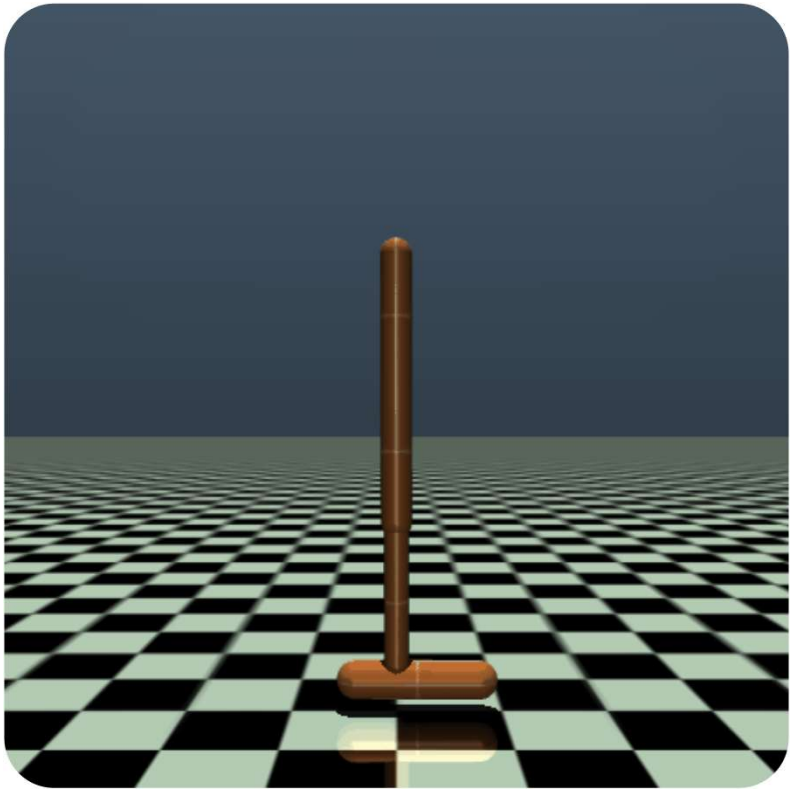


Task

- GRPO with task-success floor
- Optimized secondary objectives
- Policy π with maximum returns

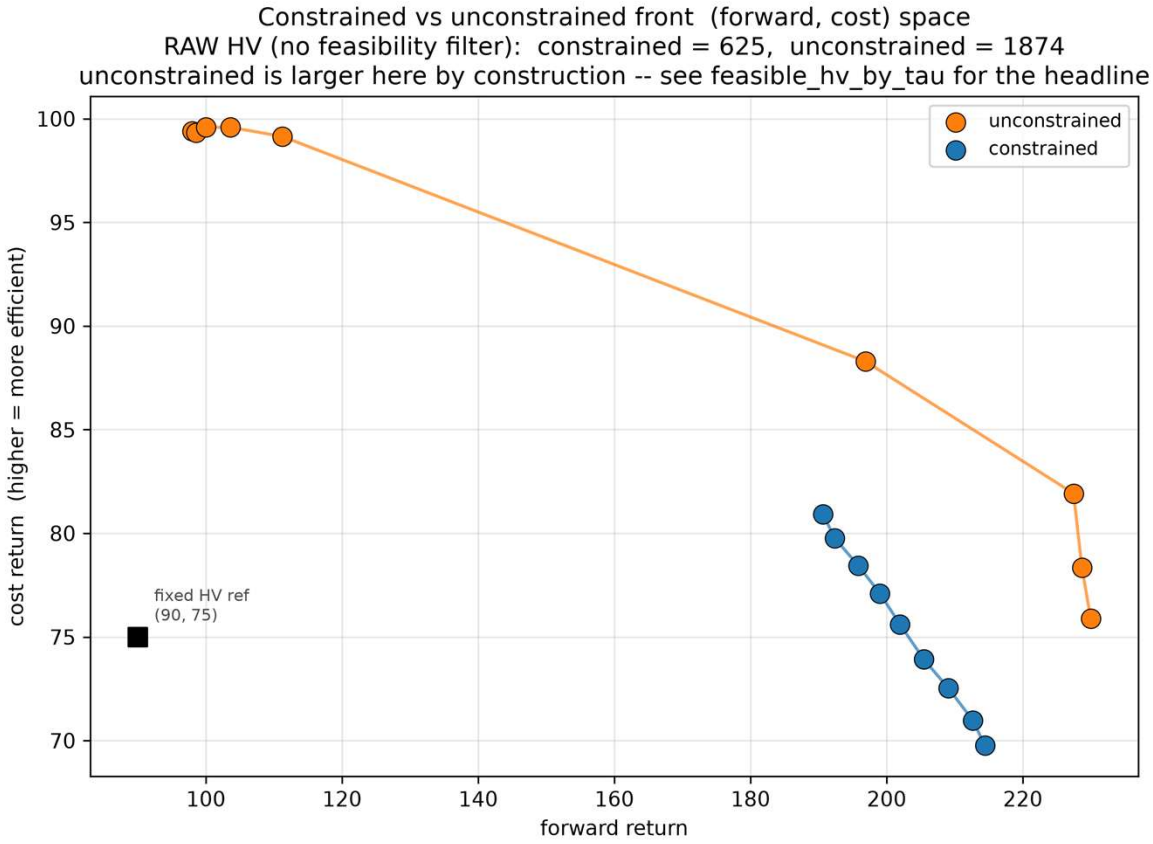
Progress: Final Environment

- “MO-Hopper”
- 3 Objectives
  - Forward (Constrained)
  - Jump Height
  - Cost
- Continuous action space
- Continuous observation space



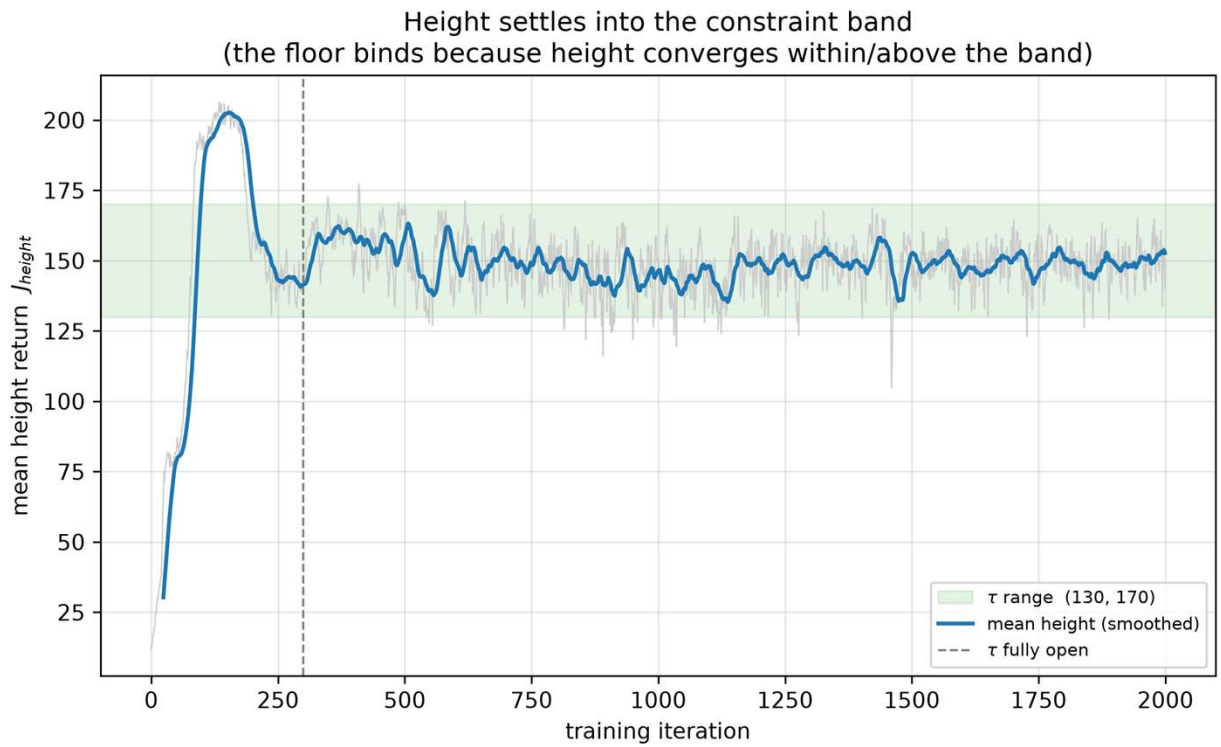
Progress: Fronts

- Forward vs cost
- Expected results

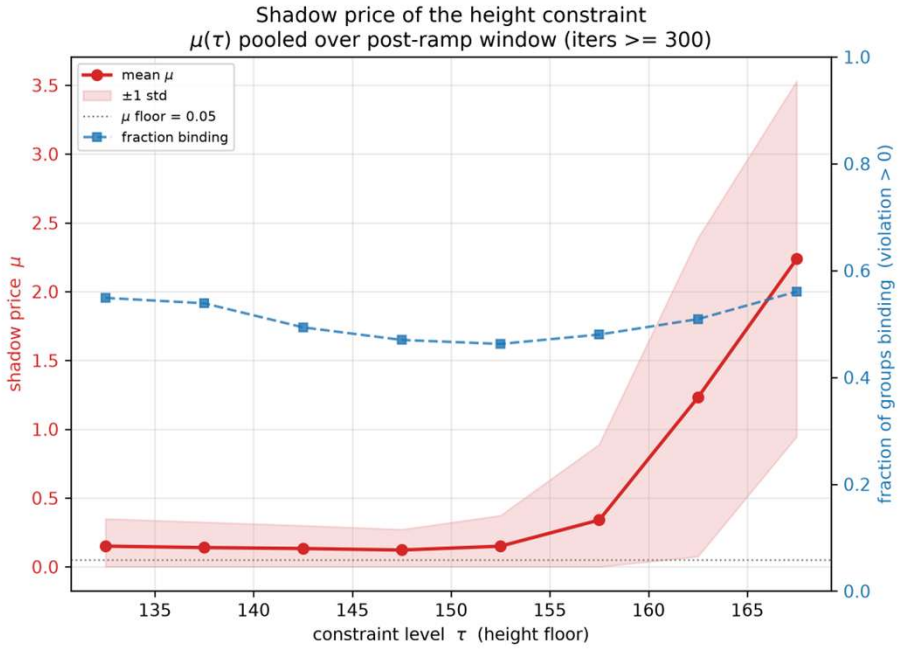


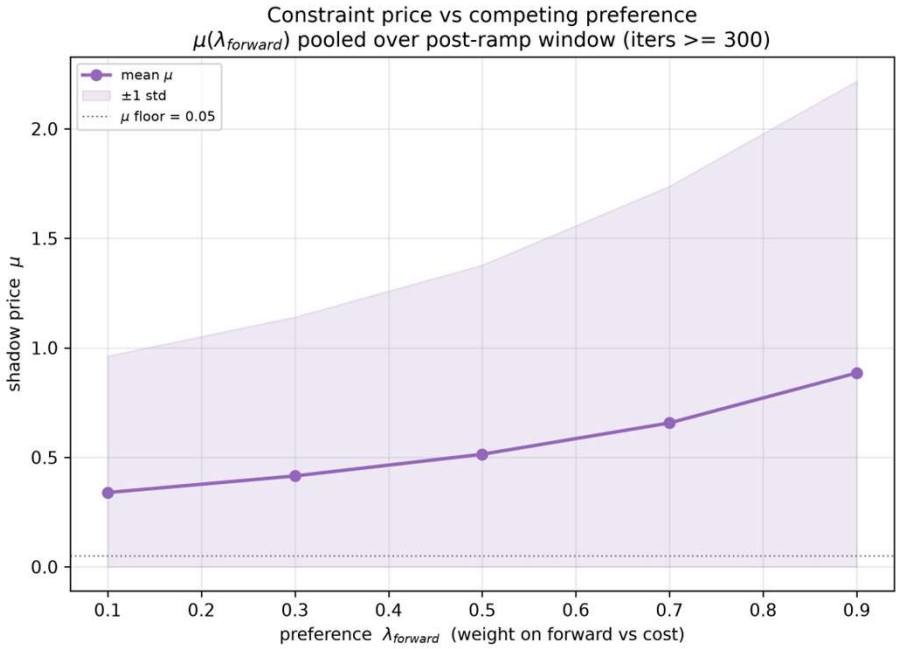
Progress: Height vs Tau

- Settles into band
- Expected results



Progress: μ vs τ and μ vs λ





- μ enforcement
- How speed affects μ

Progress: VLA Decision

- MiniVLA or OpenVLA-OFT
- **MiniVLA was selected**
  - 7x smaller
  - Libero-90 ≥ OpenVLA
  - Stanford ILIAD

Trained to 95% Action Accuracy	Libero 90
OpenVLA	62%
MiniVLA	**61.4%**



Progress: Final Environment

- Reproducing results
- Gathering success metrics



Current Work and Next Steps

Next Steps

- Finalize MiniVLA results
- Proceed with CMO-GRPO



WEEK 5 UPDATE

CODY HILL • 6/22/26