Geometric-Mean Policy Optimization

Yuzhong Zhao*, Yue Liu*, Junpeng Liu, Jingye Chen, Xun Wu, Yaru Hao, Tengchao Lv,
Shaohan Huang, Lei Cui, Qixiang Ye, Fang Wan, Furu Wei *indicates equal contribution

University of Chinese Academy of Sciences
Microsoft

Arithmetic-Mean ⇨

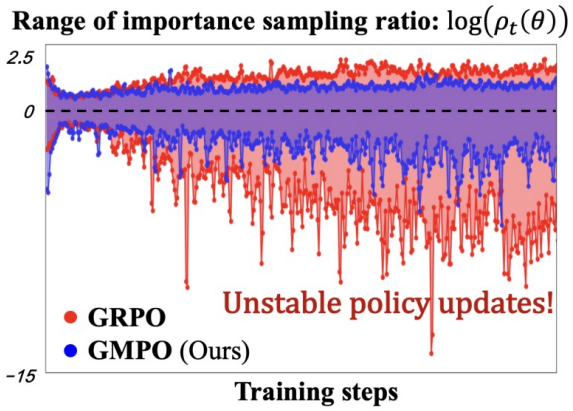
GRPO

$$\frac{1}{|o|}\sum_{t=1}^{|o|}\left\{\min\left[\rho_t(\theta)\hat{A}, \text{clip}(\rho_t(\theta), \epsilon_{\text{low}}, \epsilon_{\text{high}})\hat{A}\right]\right\}$$

Geometric-Mean ⇨

GMPO (Ours)

$$\left\{\prod_{t=1}^{|o|}\left|\min\left[\rho_t(\theta)\hat{A}, \text{clip}(\rho_t(\theta), \epsilon_{\text{low}}, \epsilon_{\text{high}})\hat{A}\right]\right|\right\}^{\frac{1}{|o|}} \cdot \text{sgn}(\hat{A})$$



Observation: GRPO(**top**) suffers from unstable policy updates due to outlier token rewards, which lead to extreme importance sampling ratios.

Solution: GMPO(**bottom**) replaces the arithmetic mean with the geometric mean to suppress outliers, resulting in more stable optimization and improved performance.

Language Model	AIME24	AMC	MATH500	Minerva	Oly.	Avg.
GRPO-1.5B Shao et al. (2024)	23.3	49.4	75.2	25.7	39.0	42.5
GMPO-1.5B (Ours)	20.0	53.0	77.6	30.1	38.7	**43.9**
GRPO-7B Shao et al. (2024)	40.0	59.0	83.4	32.4	41.3	51.2
GMPO-7B (Ours)	43.3	61.4	82.0	33.5	43.6	**52.7**
GRPO-7B Shao et al. (2024) (R1-Distill)	43.3	67.5	89.0	39.7	56.7	59.3
GMPO-7B (R1-Distill, Ours)	46.6	78.3	91.4	37.9	62.5	**63.4**

(a) Five mathematical reasoning benchmark.

Multimodal Model	Geometry3K
GRPO-7B (Shao et al., 2024)	53.3
GMPO-7B (Ours)	**54.7**

(b) Geometry3K benchmark.

MoE Model	MATH500
GRPO-32B (Shao et al., 2024)	94.6
GMPO-32B (Ours)	**96.7**

(c) MATH500 benchmark.

Agentic Model	Pick	Look	Clean	Heat	Cool	Pick2	ALL
GRPO-1.5B Shao et al. (2024)	85.3	53.7	84.5	78.2	59.7	53.5	72.8
GMPO-1.5B (Ours)	93.1	78.6	81.0	88.2	82.1	89.5	**85.9**

(d) ALFWorld benchmark.

Results: GMPO consistently outperforms GRPO across Language, Multimodal, Mixture-of-Experts, and Agentic models, demonstrating performance gains across diverse scenarios and tasks.

Video:

Code: