University of Glasgow
James Watt School of Engineering

Beyond Distributions: Geometric Action Control for Continuous Reinforcement Learning

Zhihao Lin[1]
[1] University of Glasgow
2800400L@student.gla.ac.uk

ICLR International Conference On Learning Representations

The Problem

Gaussian policies dominate continuous RL, yet their infinite support fundamentally mismatches bounded action spaces $[-1, 1]^d$. The standard tanh squashing creates gradient saturation near boundaries, affecting ≈40% of SAC training steps on HalfCheetah.

Can we eliminate this mismatch entirely?

"Actions in physical systems naturally decompose into direction and magnitude. Effective control does not require distributions; it requires geometry."

Geometric Action Control

GAC replaces distributional sampling with direct geometric operations on the unit sphere. A direction network outputs $\boldsymbol{\mu} \in \mathbb{S}^{d-1}$, and a concentration network controls exploration via learned κ:

$$\mathbf{a} = r \cdot \text{normalize}(w(\kappa) \cdot \boldsymbol{\mu} + (1 - w(\kappa)) \cdot \boldsymbol{\xi})$$

where $\boldsymbol{\xi} \sim \text{Uniform}(\mathbb{S}^{d-1})$ and $w(\kappa) = \sigma(\kappa)$. This achieves $O(d)$ complexity vs. $O(dk)$ for vMF, with 50% fewer parameters (d+1 vs. $2d$).

Code

Homepage

Architecture & Theory

GAC integrates into SAC by replacing the Gaussian policy with geometric action generation. The actor loss becomes: $L_{\text{actor}} = \mathbb{E}_s \left[\kappa(s) - \min_i Q_{\theta_i}(s, \mathbf{a}) \right]$. The learned κ acts as an endogenous exploration controller, replacing entropy regularization.

Theorem 1. The expected unnormalized sample lies along the mean direction: $\mathbb{E}_{\boldsymbol{\xi}}[\mathbf{v}] = w(\kappa)\boldsymbol{\mu}$. This provides vMF-like concentration control without Bessel function computations. As $\kappa \to \infty$, $w(\kappa) \to 1$ and samples concentrate around $\boldsymbol{\mu}$.

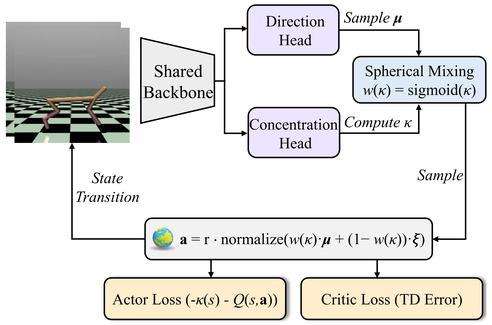

Figure 2: GAC architecture: direction head ($\boldsymbol{\mu}$) and concentration head (κ) with spherical mixing.

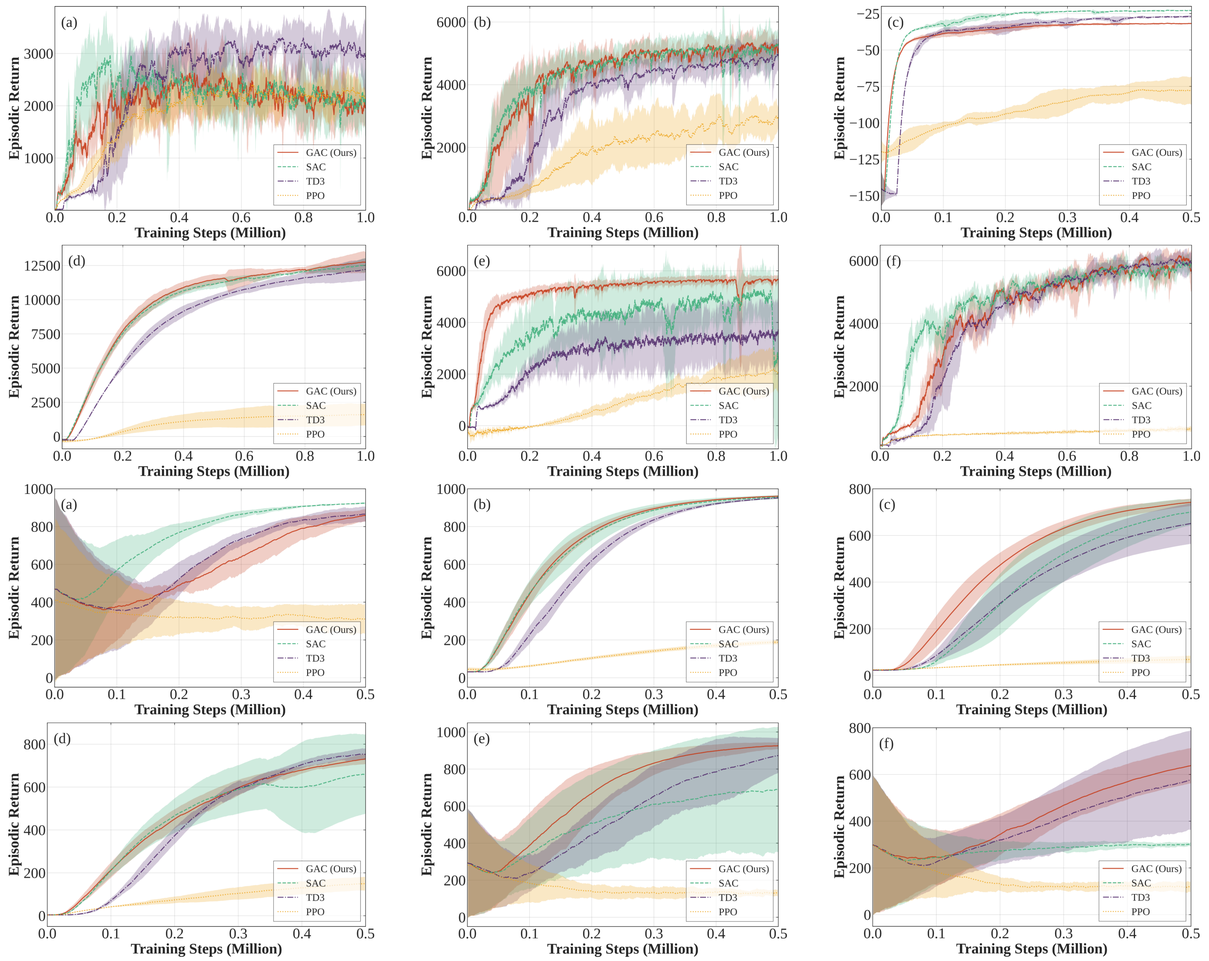

Figure 1: Learning curves on MuJoCo and DMControl benchmarks. GAC achieves best results on 9/12 tasks.

Results

GAC achieves best performance on 9 out of 12 tasks across MuJoCo and DMControl, with notable gains in high-dimensional control.

Key highlights: +37.6% over SAC on Ant-v4 (8D), +112% on quadruped-run, and competitive results on Humanoid-v4 (17D) with significantly lower variance.

Ablation & Conclusion

Ablations on HalfCheetah-v4 confirm each component is essential: removing normalization causes divergence within 5k steps, removing κ reduces performance by 10.8%.

The scaling parameter r is robust: performance varies <10% across $r \in [1.0, 3.5]$, confirming it acts as a stable geometric factor, not a fragile hyperparameter.

"Control is not about predicting densities, but about choosing directions. When geometry is respected, simplicity is not a compromise, but a strength."

Funders

This work was supported in part by the China Scholarship Council Ph.D. Scholarship for 2023-2027 (No.202206170011)