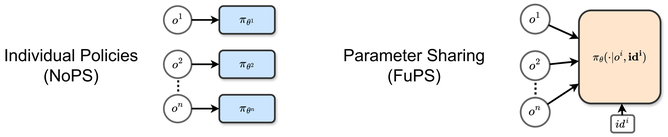
HyperMARL: Adaptive Hypernetworks for Multi-Agent RL

Kale-ab Abebe Tessera, Arrasy Rahman, Amos Storkey, Stefano V. Albrecht
k.tessera@ed.ac.uk

Paper Code

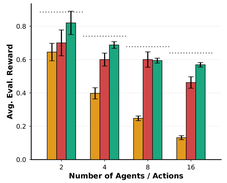
1. MOTIVATION

- **Adaptability**: Agents must flexibly learn ***specialised*** or ***homogeneous*** behaviours.
- **Trade-off**: NoPS (individual policies) is flexible but inefficient, FuPS (parameter sharing) is efficient but struggles with ***specialisation***.
- **Problem**: Efficient methods (FuPS) cannot represent ***diverse policies***.

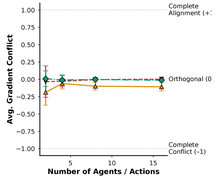


🎯**Key Insight**: **To enable specialisation, we must decouple agent-ID and observation gradients.**

Coupling them correlates with higher cross-agent gradient interference.



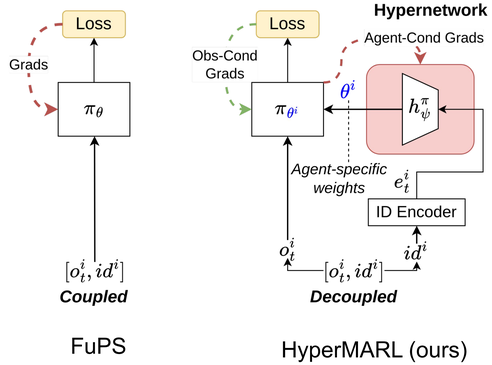
Avg. Evaluation Return



Avg. Grad Conflict

NoPS FuPS+ID FuPS+ID (No State) HyperMARL

2. 🧬 HyperMARL & Gradient Decoupling



FuPS HyperMARL (ours)

We propose ***HyperMARL***, an ***agent-conditioned hypernetwork*** that **decouples observation-** and **agent-conditioned** gradients.

Does not:
✅ change the **learning objective**.
✅ require knowing the **optimal diversity level**.
✅ require **sequential updates**.

💡***Gradient Decoupling:***

$$\nabla_\psi J(\psi) = \sum_{i=1}^{I} \underbrace{\nabla_\psi h_\psi^\pi(e^i)}_{\mathbf{J}_i \text{ (agent-conditioned)}} \underbrace{\mathbb{E}_{\mathbf{h}_t,\mathbf{a}_t\sim\boldsymbol{\pi}}\left[A(\mathbf{h}_t,\mathbf{a}_t)\,\nabla_{\theta^i}\log\pi_{\theta^i}(a_t^i \mid h_t^i)\right]}_{Z_i \text{ (observation-conditioned)}}.$$

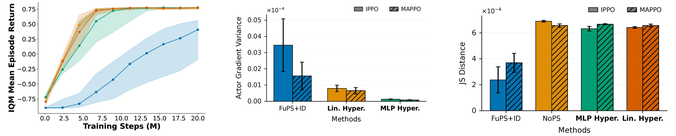
- ***Agent-conditioned:*** deterministic w.r.t mini-batch samples, separating agent identity from traj noise.
- ***Observation-conditioned***: averages trajectory noise **per agent**.

3. Results: 22 settings, up to 30 agents, 6 baselines

Across ***22 scenarios*** with up to ***30 agents***, HyperMARL performs competitively with six baselines, preserves NoPS-level behavioural diversity, and shows **lower policy-gradient variance** than FuPS.

Competitive to NoPS and Baselines in Specialised Tasks

➢ Learns **diverse** behaviours with a **lower policy variance**.

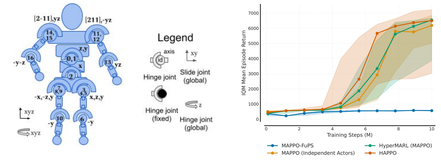


Mean Return Actor Grad. Variance Diversity

FuPS+ID NoPS **MLP Hyper.** **Lin. Hyper.**

➢ Handles **specialisation at scale** e.g. 17-agent Humanoid task.

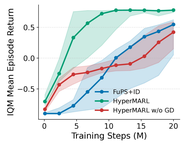
MAMuJoCo (Peng et al., 2021)



Also competitive with FuPS in homogeneous tasks. Many more results in the paper (e.g.off-policy, 30 agents).

4. Ablations + Embeddings

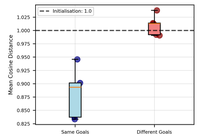
- ***Gradient decoupling matters.***



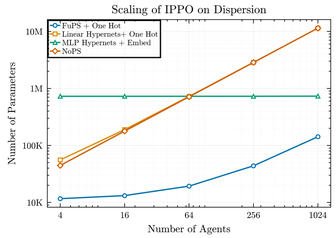
Dispersion

- ***Agent embeddings*** *move* ***closer*** *when a task requires* ***similar*** behaviour *and* ***further apart*** *when a task requires* ***different*** *behaviour.*

Navigation (Bettini et al., 2022)



5. Limitations and Scaling



- High parameter count for few agents (e.g. solution: chunked hypernets).
- **Scales efficiently** to large agent populations.

6. Main Takeaways

➢ ***Gradient decoupling*** reduces cross-agent ***gradient interference***.
➢ ***HyperMARL*** leverages this decoupling to learn ***adaptive behaviour*** without heuristics, altered objectives, or sequential updates.
➢ **Future Work**: Chunked hypernetworks & low-rank approximations.