Computing Ex Ante Equilibrium in Heterogeneous Zero-Sum Team Game

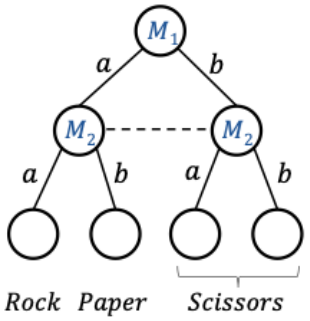
Naming LIU, Mingzhi WANG, Xihuai WANG, Weinan ZHANG, Yaodong YANG, Youzhi ZHANG, Bo AN, Ying WEN

Frontiers of Computer Science, DOI: 10.1007/s11704-025-51045-0

Problems & Ideas

- Problems of conventional Team PSRO approaches:
  - Team PSRO relies on policy sharing and therefore cannot represent the full policy space in heterogeneous team games where teammates play distinct roles.
  - This insufficient policy expressiveness prevents convergence to the global equilibrium, often leading to suboptimal solutions with high exploitability.
  - Ideas: achieves stable convergence to the global equilibrium in heterogeneous team games via heterogeneous policy parameterization and sequential correlation.

Decision Tree of Team T_1



Reward Function of Team RPS

T_1 \ T_2	(a,a)	(a,b)	(b,a)	(b,b)
(a,a)	0, 0	−1, 1	1, −1	1, −1
(a,b)	1, −1	0, 0	−1, 1	−1, 1
(b,a)	−1, 1	1, −1	0, 0	0, 0
(b,b)	−1, 1	1, −1	0, 0	0, 0

Homogeneous PSRO framework in Team RPS

	Restricted Policy Set of T_1	*Restricted Policy Set of T_2*
1	$\{(b,b)\}$	$\{(b,b)\}$
2	$\{(b,b), \overbrace{(a,a)}^{BR_{1,homo}(b,b)}\}$	$\{(b,b), \overbrace{(a,a)}^{BR_{2,homo}(b,b)}\}$
3	*Terminated with* $BR_{1,homo}(a,a) = BR_{2,homo}(a,a) = (a,a)$	

Pure strategy (a, b) of both T_1 and T_2 cannot be represented by the ***homogeneous PSRO framework***

Fig. 1 Procedure of the homogeneous PSRO framework in Team Rock-Paper-Scissors, which is a typical *heterogeneous* team game, with four agents, two teams $T_1 = \{M_1, M_2\}$ and $T_2 = \{O_1, O_2\}$, one state, and joint action spaces $\mathcal{A}_1 = \mathcal{A}_2 = \{a, b\}^2$. Agents play Rock-Paper-Scissors between the teams: if player M_1 in team T_1 (or O_1 in team T_2) chooses action b, then the team plays $Scissors$ no matter the choice of the other player in the team; if both players choose action a, then the team plays $Rock$; otherwise, the team plays $Paper$. The two players in team T_1 or opponent team T_2 are *heterogeneous* because the actions a and b serve different functions for them. Specifically, player M_1 (or O_1) can unilaterally choose the team decision $Scissors$ by playing action b, while player M_2 (or O_2) must coordinate with the other player to choose $Paper$ by playing action b.

Main Contributions

- Contributions:
  - We identify and theoretically characterize the fundamental limitations of the homogeneous PSRO framework in heterogeneous team games;
  - A novel sequential correlation mechanism that optimizes heterogeneous players' policies in a serialized manner, avoiding the exponential growth of joint optimization policy spaces;
  - The first PSRO framework for heterogeneous team games, which integrates the sequential mechanism into an iterative procedure.

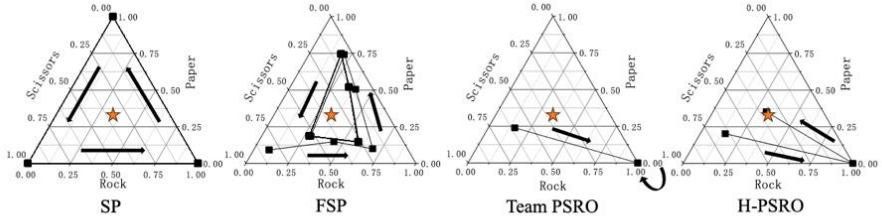


Fig. 1 Trajectories of SP, FSP, Team PSRO and H-PSRO in Team RPS game. H-PSRO shows superior convergence to the global TMECor due to the sufficient policy expressiveness of heterogeneous policies and the corresponding full equilibrium expressiveness under the heterogeneous PSRO framework (see Theorem 1).

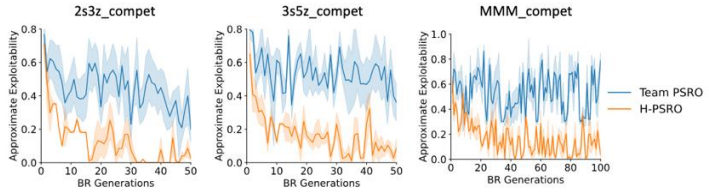
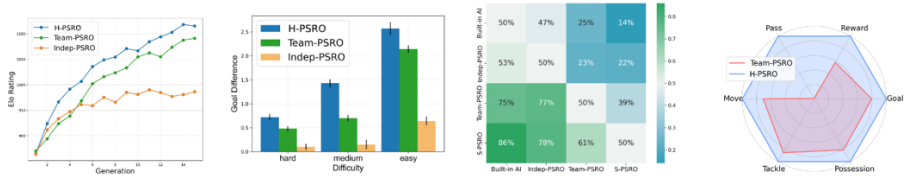


Fig. 2 Exploitability of H-PSRO and Team PSRO in the Competitive StarCraft Benchmark.



(a) Elo Rating. (b) Goal Difference. (c) Relative PP. (d) Performance Radar.

Fig. 3 Performance of H-PSRO, Team PSRO and Indep-PSRO in Google Research Football.