**Experiment Plan for Exploring the Tournament Equilibrium**

**I Methodology**

1. Gradient-based optimization: Directly optimize two variables, the efforts of player 1 and player 2, using gradient-based methods. A benchmark solution is available in closed form, making it suitable for evaluating numerical convergence.

This approach is limited to static, symmetric, and single-round games. It does not generalize well to more complex settings involving multi-stage interactions, more than two players, additional randomness, learning from trajectory (sequence of states and actions).

1. Reinforcement learning with self-play:

We adopt a reinforcement learning framework where both agents are trained via self-play using policy-based methods. Two optimization techniques are considered:

1）Policy gradient-REINFORCE

* Objective function:

Where:

: Parameters of Policy network;

= (s1,a1, …, sT, aT): Trajectory of states and actions;

: total reward along the trajectory.

* Gradient of the objective:

2）Proximal Policy Optimization (PPO)

* Clipped surrogate objective function: More stable and sample-efficient training procedure by preventing large, destructive updates to the policy

Where:

, probability ratio between new and old policies

: the estimation of the advantage,

: clip threshold to constrain policy updates for stability.

**II Training with different parameters**

|  |  |
| --- | --- |
| 1 | wH = 6.5, wL = 3 |
| 2 | k = 1/2500 = 0.0004 |
| 3 | Efforts : [0,100], [0,200]; |
| 4 | Noise term : ; ; ; |

**III Three groups of experiments**

1. **One-stage tournament with two identical players**

|  |  |
| --- | --- |
| Utility function |  |
| Output of stage one | + , with , |
| Cost of effort |  |
| The player’s expected utility |  |
| The symmetric equilibrium effort |  |

1. **One-stage Tournament Expansion** 
   1. Three identical competitors: One winner, two losers

|  |  |
| --- | --- |
| Output of stage one | + , with , |
| Cost of effort |  |
| The player’s expected utility |  |
| The symmetric equilibrium effort |  |

* 1. Two different competitors
     1. Different cost function: ,

|  |  |
| --- | --- |
| Output of stage one | + , with , |
| Cost of effort |  |
| The player’s expected utility |  |
| The symmetric equilibrium effort |  |

:

:

* + 1. Different ability: + , ,

|  |  |
| --- | --- |
| Output of stage one | + , with , |
| Cost of effort |  |
| The player’s expected payoff |  |
| The symmetric equilibrium effort |  |

1. **Two-stage tournament**
2. Utility function:
3. Output of stage t: + , with
4. Cost of effort :
5. The player’s expected payoff is
6. The symmetric equilibrium effort in stage 1: ; The symmetric expected equilibrium effort in stage 2:

IV Experiment analysis

1. Exploring the equilibrium: Evaluate whether the strategies learned through different methods (Gradient-based optimization, REINFORCE, PPO) converge to the theoretically derived Nash equilibrium in the one-stage tournament. The efforts selected by agents will be compared against the closed-form analytical solution under various parameter settings.
2. Generalization performance of Reinforcement learning
3. Stability and convergence Comparison: Compare the stability and convergence behavior of the three optimizations across different parameter setups.