Multi-Agent MIMO Strategies for Coordination in Distributed Wireless Networks

Minning Liu, Emmanuelle Padilla

Problem & Baseline

The Problem: Distributed MIMO systems suffer from severe interference when the agents don't coordinate

Our Baseline is Independent Q-Learning (IQL) because it's a non-cooperative approach. We chose IQL because its primary limitation is agents treat one another as part of the environment which leads to cyclic loops or unstable convergence. (Basically we do not achieve a satisfactory equilibria → we have maximized our rewards and this is the best policy moving forward)

Goal: Beat this baseline using the SeqComm algorithm

Input Data

- 3 wireless agents
- Each agent has a 4-antenna transmitter and a 2-antenna receiver
- Complex Rayleigh fading channels
- A Gauss-Markov AR(1) update to simulate realistic mobile fading
- Interference calculations between all agents
- Noise floor of –100 dBm

Coding Progress

Our attempt at a IQL Baseline with wireless physics:

AR1 Channel Modeling → simulates memory fading over time

DFT Beamforming → uses complex vectors to represent a realistic 5G.

Target Networks → prevents the "moving target" and stabilizes our physics simulation

These functions create a test data/environment and act like a template for actual data like DeepMIMO

Algorithms

For algorithms, we use Independent Q-Learning (IQL) for each wireless agent.

Each agent learns two things:

1. A beamforming vector chosen from a DFT codebook.
2. A transmit power level.

Each agent receives a reward based on: $\text{Rate} = \log 2(1 + \text{SINR}) - 0.05 \times \text{Power}$

The goal is to increase throughput while reducing interference.

Metrics & Evaluation

We evaluate performance using standard wireless communication metrics:

- SINR (Signal-to-Interference-plus-Noise Ratio)
- Total episode reward, which represents system throughput
- Interference level between agents

The primary metric will be Spectral Efficiency (Sum Rate)

- Formula $R = \sum \log_2(1+SINR)$
  - X-Axis: Time Step
  - Y=Axis: Sum Rate (Throughput)

What We Have Achieved

The IQL baseline runs and learns, it:

- Generate realistic 4x2 MIMO channels
- Compute SINR and throughput
- Calculate interference between all agents
- Train 3 agents using Independent Q-Learning
- Update the wireless fading over time
- Learn beamforming and power strategies that increase reward

```
MIMO Env: 3 Agents, 4x2 Antennas, Action Space: 40
Episode 50 | Avg Reward: -521.68 | Epsilon: 0.778
Episode 100 | Avg Reward: -386.93 | Epsilon: 0.606
Episode 150 | Avg Reward: -248.28 | Epsilon: 0.471
Episode 200 | Avg Reward: -215.31 | Epsilon: 0.367
Episode 250 | Avg Reward: -177.80 | Epsilon: 0.286
Episode 300 | Avg Reward: -133.59 | Epsilon: 0.222
Episode 350 | Avg Reward: -117.99 | Epsilon: 0.173
Episode 400 | Avg Reward: -81.76 | Epsilon: 0.135
Episode 450 | Avg Reward: -61.91 | Epsilon: 0.105
Episode 500 | Avg Reward: -57.32 | Epsilon: 0.082
Episode 550 | Avg Reward: -31.46 | Epsilon: 0.063
Episode 600 | Avg Reward: -22.07 | Epsilon: 0.050
Episode 650 | Avg Reward: -16.31 | Epsilon: 0.050
Episode 700 | Avg Reward: -13.98 | Epsilon: 0.050
Episode 750 | Avg Reward: 8.05 | Epsilon: 0.050
Episode 800 | Avg Reward: -4.96 | Epsilon: 0.050
Episode 850 | Avg Reward: -2.58 | Epsilon: 0.050
Episode 900 | Avg Reward: 3.16 | Epsilon: 0.050
Episode 950 | Avg Reward: 2.76 | Epsilon: 0.050
Episode 1000 | Avg Reward: -6.54 | Epsilon: 0.050
```

What’s Next

SeqComm Implementation

- Goal: Upgrade the current IQL agents to communicate before acting
- Action: Add a "Communication Round" to our Python environment where agents exchange "Intention Values" (estimated rewards) to decide who transmits first

Experiments & Validation Goal: Prove our hypothesis that coordination beats independent learning.

- Run head-to-head simulations:
  - Scenario A: Independent Q-Learning
  - Scenario B: SeqComm
- Generate the final Spectral Efficiency vs. Episode graph