Case Study - Generating Mixed Strategy Profiles from Correlated Equilibrium in Arbitrary Size Two-Player Games

Alysson Ribeiro da Silva

I. Introduction

This report presents a solution for a Multi-agent Exploration [1], [2] problem based on a set of Mixed Strategies heuristically approximated from a Correlated Equilibrium generated through a Probability Density Function over viable outcomes. A video of each experiment is available[1][2][3]. Furthermore, this case study presents one optimization problem to find a correlated equilibrium using a distribution over a set of viable outcomes. Consider the following assumption.

Assumption 1. *All robots can reach all frontiers in the M_t.*

The aforementioned assumption was made to facilitate the description of the utility of all outcomes for all robots. Consider the following problem regarding the Multi-agent SLAM with Frontier Exploration.

Problem 1 (Two robots and two zones exploration strategies)**.** *Consider the existence of two zones, Zone 1 and Zone 2, and their respective information gains u'_1 and u'_2. For brevity, Zone 1 will be called only by the number 1 and Zone 2 will be called only by its number 2. Consider the existence of two robots that need to solve one instance of a Markov Game, where the utility $\sum u'_i(a)$ of an outcome for robot i is the gain in information for all frontiers $f^* \in F$ of the zone being visited. Consider that if both robots select the same zone to explore, their utility will be computed in a similar fashion to a Congestion Game, or $u_i \leftarrow -u_i$. Let $N = 2$, $A_1 = (1, 2)$, $A_2 = (1, 2)$, $O = A$. Find the best-mixed strategies Π_1 and Π_2 that approximate the correlated equilibrium of the game to a probability density function represented by Υ and maximizes the total expected payoff of the game in the presence of an environment representation M_t at the time t.*

As a benchmark for this case study, consider the game described in Table I as an instance of Problem 1.

		Robot 2	
		zone 1	zone 2
Robot 1	zone 1	-2,-2	5, 5
	zone 2	6, 5	-3,-3

TABLE I: Two robots game.

where, $A_1 = (1, 2)$ and $A_2 = (1, 2)$. Set Υ as a logarithmic curve surrounding the best outcome $(zone2, zone1)$ that spreads to $(zone1, zone2)$. Consider the outcomes $(zone1, zone1)$ and $(zone2, zone2)$ to be conflicting visitation areas. Solve the following optimization problem.

$$
\begin{aligned}
&\text{maximize:} \\
&\quad f(.) = \omega_1(-4\Pi_1(a_1)\Pi_2(a_1) + 10\Pi_1(a_1)\Pi_2(a_2) \\
&\qquad\qquad + 11\Pi_1(a_2)\Pi_2(a_1) - 6\Pi_1(a_2)\Pi_2(a_2)) \\
&\qquad\qquad + \omega_2\zeta(\Lambda.\Upsilon + 1) \\
&\text{subject to:} \quad \sum \Pi_i(a) = 1, \forall i \in N \\
&\qquad\qquad\quad \sum \upsilon_i = 1, \upsilon_i \in \Upsilon, \text{ where } G(\upsilon_i) = 1 \\
&\qquad\qquad\quad \Pi(a) \geq 0
\end{aligned}
$$

A. Experimental Configuration

To solve the aforementioned maximization instance of Problem 1, the mixed strategy profiles and variables inside Λ were coded as learnable parameters in a Genetic Algorithm (GA) with the PyGAD python package. Four different optimization procedures were performed to obtain different profiles. Furthermore, two experiments were conducted, where the first approach $\Pi_1\Pi_2$ to Λ, where the density of the $N(\mu, \sigma^2)$ is fully centered around μ with low . Differently, the second experiment uses a higher σ in Λ to achieve exploration-inclined mixed profiles around the best outcome in the game.

After the optimization process, a mixed strategy profile matrix was also extracted from the obtained solution as a function of Π_1 and Π_2. The best fitness value for the performed optimization was extracted for each generation of the GA. The GA runs for 1000 generations with a population size of 500 and 10 breeding individuals between generations with uniform crossover. A random mutation scheme was set to 0.5% for each gene and all variables were set to work inside the $[0, 1]$ interval.

B. Low Variance Numerical Results

As shown in Tables II and III, the optimization generated similar mixed strategies for all 4 tests. In contrast, the correlated equilibrium matrices seem to have converged into the desired outcome. The behavior portrayed could have happened since Υ was set with a high variance surrounding the best possible outcome. In contrast, the mixed strategies found that exploring zone 1 and zone 2 is the best possible action profile.

As shown in Fig 1, the optimization converged for all evaluations to a maximum fitness near 2 which indicates

[1] no equilibrium: https://youtu.be/0xqYxiQCdbA
[2] equilibrium short: https://youtu.be/i3V1tRCEIRQ
[3] equilibrium long: https://youtu.be/sF5uBQC0rJ4

1	zone 1	zone 2
robot 1	0.0	0.98
robot 2	0.99	0.002

2	zone 1	zone 2
robot 1	0.01	0.98
robot 2	0.99	0.003

3	zone 1	zone 2
robot 1	0.0	0.99
robot 2	0.99	0.0

4	zone 1	zone 2
robot 1	0.0	0.98
robot 2	0.99	0.0

TABLE II: Low variance Λ generated mixed strategies.

1	zone 1	zone 2
zone 1	0.01	0.0
zone 2	0.98	0.002

2	zone 1	zone 2
zone 1	0.01	0.0
zone 2	0.98	0.003

3	zone 1	zone 2
zone 1	0.002	0.0
zone 2	0.99	0.0

4	zone 1	zone 2
zone 1	0.01	0.0
zone 2	0.98	0.0

TABLE III: Extracted correlated equilibrium.

the maximum payoff that can be obtained by approximating $\Pi_1\Pi_2$ to Λ. The agent converged near generation 200 for all tests, which may indicate that the global maximum of the objective function is easily found due to its slope. This behavior may help to generate solutions in the real world for the same problem formulation in an efficient manner.

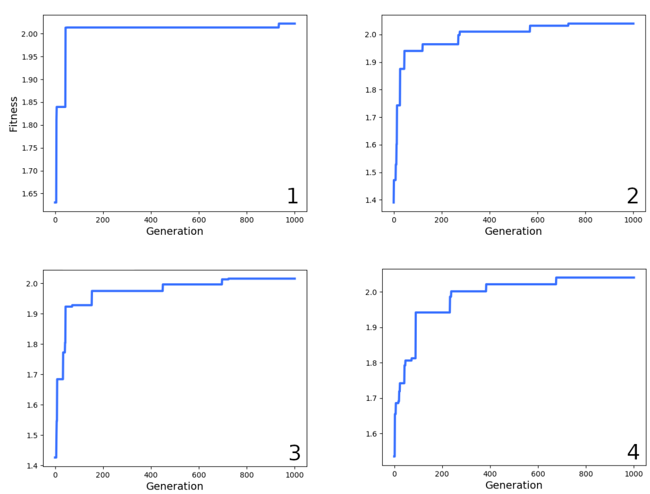

Fig. 1: Fitness convergence of the best individual for all 4 evaluations with low variance. The x-axis represents each generation and the y-axis is the fitness for the best individual from the GA.

C. High Variance Numerical Results

One of the main benefits of the proposed method is the fact that it does not necessarily converge towards the best possible equilibrium with a probability of 1. As depicted in Tables IV and V, when increasing the variance of Λ, the agents will explore different outcomes. The results show the robots could explore the same zone. From the point of view of the problem formulation, this is not the desired behavior since they will share the same resources. However, it should be mentioned that the aforementioned behavior happened specifically for the proposed problem instance. In reality, in a game with many possible exploration zones, it can be beneficial to be inclined to explore viable outcomes surrounding the best one due to local minima and uncertainty.

1	zone 1	zone 2
robot 1	0.26	0.73
robot 2	0.73	0.26

2	zone 1	zone 2
robot 1	0.26	0.73
robot 2	0.73	0.26

3	zone 1	zone 2
robot 1	0.17	0.82
robot 2	0.83	0.16

4	zone 1	zone 2
robot 1	0.27	0.72
robot 2	0.71	0.28

TABLE IV: High variance Λ generated mixed strategies.

1	zone 1	zone 2
zone 1	0.19	0.06
zone 2	0.54	0.19

2	zone 1	zone 2
zone 1	0.19	0.07
zone 2	0.53	0.19

3	zone 1	zone 2
zone 1	0.14	0.02
zone 2	0.68	0.13

4	zone 1	zone 2
zone 1	0.19	0.08
zone 2	0.51	0.2

TABLE V: Extracted correlated equilibrium.

The high variance fitness convergence is presented in Fig 2. According to the observed optimization behavior, the agent was able to generate viable solutions that explore its surroundings and reach a maximum fitness near 0.7. The obtained fitness is smaller than the ones obtained for a low variance since the robot will explore solutions that do not necessarily maximize the overall payoff of the game.

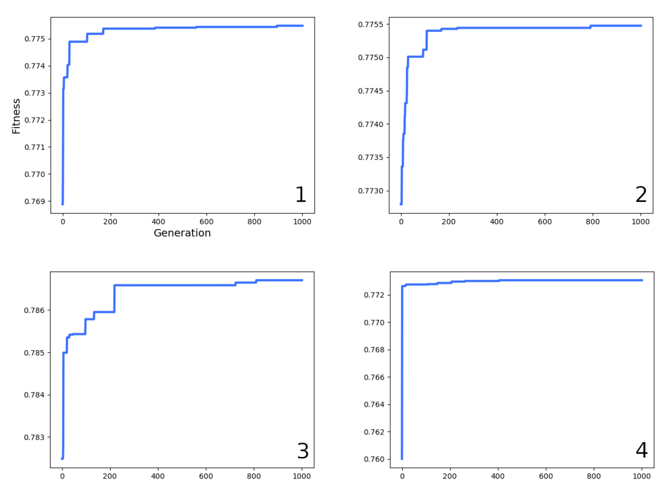

Fig. 2: Fitness convergence of the best individual for all 4 evaluations with high variance. The x-axis represents each generation and the y-axis is the fitness for the best individual from the GA.

D. Conclusion

In this case study, it was presented a normal form game formulated with the method described in Section **??** and used to generate a set of mixed strategy profiles. The generation of the strategies was achieved following the desired equilibrium described by a normal distribution as a function of all mixed strategy profiles. A solution represents an exploration zone strategy for all robots that can be computed when their communication is possible. During all evaluations, the agent was able to find feasible solutions through a GA (Genetic Algorithm) for low and high variances applied over the desired distribution. Despite the fact that it reached smaller fitness values for the high variance approximation, it portrayed a convergence time similar to the one depicted by the low variance model. Since the convergence time seems to be the same regardless of the variance from Λ, it is possible that a robot could find any feasible solution in agile time.

Among its advantages, if compared to classic correlated equilibrium approaches, the presented method allows to approximate an equilibrium as a function of mixed strategy profiles, which is the desired behavior for real-world deployment due to the nature of the Multi-agent SLAM problem. Furthermore, the formulation allows the problem to always have a solution regardless of the existence of dominant strategies since it approximates a correlated equilibrium to any arbitrary PDF (Probability Density Function).

REFERENCES

[1] M. U. M. Bhutta, M. Kuse, R. Fan, Y. Liu, and M. Liu, "Loop-box: Multiagent direct slam triggered by single loop closure for large-scale mapping," *IEEE Transactions on Cybernetics*, vol. 52, no. 6, pp. 5088–5097, 2022.

[2] M. Karrer, P. Schmuck, and M. Chli, "Cvi-slam—collaborative visual-inertial slam," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 2762–2769, 2018.