Evolving Mixed Nash Equilibria for Bimatrix Games

David Iclănzan
david.iclanzan@gmail.com

Gaskó Noémi
gaskonomi@yahoo.com

D. Dumitrescu
ddumitr@cs.ubbcluj.ro

Babeş-Bolyai University, Kogalniceanu no. 1, Cluj-Napoca, 400084, Romania

ABSTRACT

In a mixed strategy equilibrium players randomize between their actions according to a very specific probability distribution, even though with regard to the game payoff, they are indifferent between their actions. Currently, there is no compelling model explaining why and how agents may randomize their decisions is such a way, in real world scenarios.

In this paper we experiment with a model for bimatrix games, where the goal of the players is to find robust strategies for which the uncertainty in the outcome of the opponent is reduced as much as possible. We show that in an evolutionary setting, the proposed model converges to mixed strategy profiles, if these exist. The result suggest that only local knowledge of the game is sufficient to attain the adaptive convergence.

Categories and Subject Descriptors

I.2.8 [**Artificial Intelligence**]: Problem Solving, Control Methods and Search

General Terms

Algorithms, Design, Theory

1. INTRODUCTION

In Game Theory a mixed strategy is a probability distribution over the actions available for a player. If only one action has a positive probability (of one) to be selected, the player is said to use a pure strategy. A mixed strategy profile induces a probability distribution over the outcomes of the game.

A Nash equilibrium is a strategy profile with the property that no single player can gain an advantage by deviating unilaterally to another strategy. Every game with a finite set of players and actions has at least one such equilibria.

The concept of mixed strategies is a fundamental component of game theory, as it can provide Nash equilibria in games where no equilibrium in pure strategies exists. However, the empirical relevance of mixed strategies have been often criticized for being "intuitively problematic" [1]. Albeit there are theoretical arguments trying to rationalize this concept [3], it is not clear why and how players randomize their decisions. Beside the behavioral observation that people seldom make their choices following a lottery, the most puzzling question arises from the "indifference" property of a mixed strategy equilibrium. In mixed equilibrium, given the strategies chosen by the other players, each player is indifferent among all the actions that he may select with positive probability, as they do not affect the resulting payoff. Therefore, there is no direct benefit to select precisely the strategy that induces the opponents to be indifferent, as required for the existence of the equilibrium. Then, in the absence of communication between players, how can a mixed equilibrium arise in a real-world scenario, especially in cases of incomplete information?

In this paper we experiment with a novel model, that can lead to the emergence of mixed equilibrium. Here, agents aim to develop strategies for which the payoff outcome of the opponent can be predicted.

2. PROPOSED MODEL

Rational agents often build internal models that anticipate the *actions* of the other players and adapt their strategy accordingly. Here, we experiment with a model for bimatrix games where players try to anticipate directly the *game payoff* of the other player. The agents adapt their strategy in order to reduce the uncertainty of this prediction. This is in contrast with the classical scenario, where players foremost objective is to maximize their game utility.

Let (w, p) be a mixed strategy profile for a bimatrix game, where w defines the probability distribution over the actions available for the first player, with p having a similar role for the second player. Let $u_1(w, p)$ and $u_2(w, p)$ denote the game payoff for player one respectively player two.

Then, the proposed model is formalized as follows:

$$\begin{cases} o_1 = \underset{w}{\operatorname{argmin}}\left(\frac{1}{n}\sum_{i=1}^{n}(u_2(w,p) - u_2(w, \delta_i(p)))^2\right) \\ o_2 = \underset{p}{\operatorname{argmin}}\left(\frac{1}{n}\sum_{i=1}^{n}(u_1(w,p) - u_1(\delta_i(w), p))^2\right) \end{cases} \quad (1)$$

where δ_i provides a perturbation to the input probability distribution, and n is the number of perturbations.

If a mixed strategy equilibrium exits, it will optimally satisfy this multiobjective model, as a direct consequence of the "indifference" property, with both objective values equaling 0. Furthermore, the model provides incentive to not deviate from this strategy profile. If an agent deviates from the equilibrium, the other player is not completely indifferent, the squared average difference of different plays might deviate from the zero minimum.

Several important questions arise regarding the proposed model:

GECCO'12 Companion, July 7–11, 2012, Philadelphia, PA, USA.
ACM 978-1-4503-1178-6/12/07.

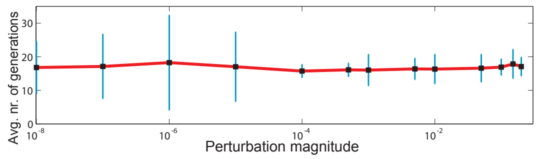

Figure 1: Semilogarithmic plot of the average convergence speed for various perturbation magnitudes.

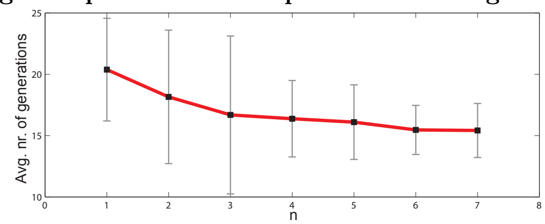

Figure 2: Average convergence speed when using various number of perturbed states.

- Can the model be used to locate a mixed strategy equilibrium (if it exists), using an adaptive search?
- What should the magnitudes of perturbations provided by δ be? Arbitrary perturbation equates to a complete information game, where one can internally evaluate every possible strategy profile, while small perturbations assume only a local knowledge about the game outcomes.
- How many perturbations of the actual strategy profile are required at each step for reliable results (how big should parameter n be?

We empirically investigate these issues in the next section.

3. DETECTION METHOD AND RESULTS

The proposed model is optimized using the NSGA II [2] – a fast multiobjective evolutive algorithm based on the concept of nondomination.

NSGA II segregates the population into layers, according to their domination degree and inside each layer, a diversity enhancing sharing function is employed to assign fitnesses. The elitist selection takes into account both the rank and the diversity maintaining crowding distance factor.

In our setup, an initial population of 60 individuals are generated randomly, where each individual encodes a strategy profile. At each step the nondominated individuals from the actual population can be considered as the approximation of certain equilibria. As genetic operators, crossover and mutation for real values are used, with probability 0.8 and 0.01. For the test problems, selection, recombination, and mutation is repeated in the bound of 500 generations. A run is considered successful, if a strategy profile is located that is very close to the target equilibrium state i.e. the euclidean distance between the two points is less or equal then a preset threshold $\epsilon = 0.0001$.

While experiments have been conducted for several games, due space considerations we present the data for only one case. So far, results for the other experiments are inline with the conclusions drawn from this case.

In the studied game each player can choose from two action. The payoff matrix is:

		Player 2	
	Strategy	p	1-p
Player 1	w	(9,9)	(6,10)
	1-w	(10,6)	(0,0)

Table 1: Game with two pure Nash equilibria and one mixed equilibrium at $(6/7, 6/7)$

The game has two pure Nash equilibria points and one mixed equilibrium at $(6/7, 6/7)$.

In a first experiment, we set parameter n to a constant value of 5 and studied the effect of perturbation magnitude on the convergence of the model. In the analyzed approach, a player strategy profile is perturbed by adding a gaussian noise with a standard deviation of σ, where σ took the following values: [0.00000001, 0.0000001, 0.000001, 0.00001, 0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1, 0.15, 0.2]. For each σ 500 independent runs were performed and for the successful runs the average convergence generation and standard deviation was computed. From the total of 6500 runs only in 9 cases (0.0014%) the mixed equilibrium was not located with sufficient exactness. The obtained average is displayed in Fig. 1. Surprisingly, as one can see, the search is mostly insensitive to the amount of perturbation.

In a next step, we locked $\sigma = 0.00000001$ and experimented with various number of perturbations (parameter n) used at each evaluation, ranging from 1 to 7. Again, an average of 500 independent run for each case was computed. The result is presented in Fig. 2. The model is moderately sensitive to the parameter n, the lower is this number, the higher is the required average number of generations until convergence. The difference between using only one perturbed point and using seven points is on average 4.9627 generations, representing an 32.19% increase.

4. CONCLUSIONS

We proposed a model that adaptively converge to mixed strategy equilibria, when optimized via an evolutionary multiobjective search method. The model can work with only a local knowledge about the game, centered around the actual strategy profile, and at each step requires only one evaluation of a slightly perturbed strategy profile. Future work will extend the model to more than two players.

Acknowledgments
David Iclănzan acknowledges the financial support of the Sectoral Operational Program for Human Resources Development 2007-2013, co-financed by the European Social Fund, within the project POSDRU 89/1.5/S/60189 with the title "Postdoctoral Programs for Sustainable Development in a Knowledge Based Society".

5. REFERENCES

[1] R.J. Aumann. What is game theory trying to accomplish? In *Frontiers of Economics, edited by K. Arrow and S. Honkapohja*. Citeseer, 1985.

[2] Kalyanmoy Deb et al. A fast elitist non-dominated sorting genetic algorithm for multi-objective optimisation: NSGA-II. In Marc Schoenauer et al., editor, *PPSN VI*, volume 1917 of *LNCS*, pages 849–858, Paris, France, 16-20 Sept 2000. Springer-Verlag.

[3] J.C. Harsanyi. Games with randomly disturbed payoffs: A new rationale for mixed-strategy equilibrium points. *International Journal of Game Theory*, 2(1):1–23, 1973.