Beating the Best Nash without Regret

Katrina Ligett
Cornell University, Department of Computer Science
and
Georgios Piliouras
Georgia Institute of Technology, School of Electrical and Computer Engineering
Johns Hopkins University, Department of Economics

1. INTRODUCTION

Nash equilibrium analysis has become the *de facto* solution standard in game theory. This approach, despite its prominent role, has been the subject of much criticism for being too *optimistic*. Indeed, in general games, natural play need not converge to Nash equilibria. In games with multiple equilibria, it is unclear how players are expected to coordinate; even in games with a unique equilibrium, finding it may involve unreasonable expectations on player communication or computation.

Nevertheless, Nash equilibrium analysis has taken much of this criticism in stride. After all, to paraphrase von Neumann, the truth is too complicated to allow anything but approximations. In the case of Nash equilibria, one would hope they make accurate predictions about useful measures of the dynamic learning behavior of the agents. One typical such measure of interest is the social welfare, defined as the sum of the utilities of the agents. A significant and diverse volume of theoretical work (including work on generalizations of the price of anarchy and on equilibrium selection) has gone into showing that Nash equilibrium analysis indeed works well as an optimistic measure of performance, providing useful upper bounds.

By contrast, in this note, we discuss a recent stream of work that shows that selfish adaptive play can in some settings achieve *arbitrarily higher* social welfare than even the *best* Nash equilibrium. This work models selfish adaptive play in a repeated game by supposing that agents employ simple learning algorithms to select their actions. In particular, these algorithms are drawn from the learning paradigm in which all agents minimize their long term *regret*.

An agent's regret in a repeated game is a measure of, in hindsight, how much her performance could have been improved by instead selecting the best single fixed action over all game rounds. There exist simple, natural algorithms that achieve low regret (though the definition itself is not tied to any specific algorithm, and merely reflects successful long-term behavior). Although worst-case regret-minimizing algorithms cannot beat even the worst Nash equilibrium, as we will see, natural regret-minimizing learning dynamics can sometimes beat not only the worst Nash equilibrium, but even the best.

Authors' Addresses: Katrina Ligett, Upson Hall, Computer Science Department, Cornell University, Ithaca NY 14853. Email:katrina@cs.cornell.edu
Georgios Piliouras, Technology Square Research Building, Georgia Institute of Technology, Atlanta GA 30332. Email:georgios.piliouras@gmail.com

2. OPTIMAL CYCLIC ATTRACTORS WITHOUT REGRET

In [Kleinberg et al. 2011], we (along with R. Kleinberg and E. Tardos) consider a stylized game where cyclic phenomena arise naturally. The game is constructed to provide a proof of concept: even in a simple game with a unique Nash equilibrium, simple learning dynamics may outperform that equilibrium. We show that natural regret-minimizing algorithms converge to cyclic attractors that exhibit optimal social welfare, which can be arbitrarily better than the best Nash equilibrium, even in games of constant size (a constant number of agents and strategies).

The game we consider is an uneven variant of matching pennies played along the edges of a cycle on the agents. We call this game Asymmetric Cyclic Matching Pennies. There are three agents numbered $1, 2, 3$, with two strategies each, H and T. The utility of agent i depends only on his action and the action of agent $i-1$, as shown in Figure 1. If agent i's strategy matches the strategy of agent $i-1$, then i receives 0 payoff.[1] If agent i plays strategy H whereas agent $i-1$ plays strategy T, then i receives a payoff of 1. Lastly, if agent i plays strategy T whereas agent $i-1$ plays strategy H, then i receives a payoff of $M \geq 1$.

agent i \ agent $i-1$	H	T
H	0	1
T	M	0

Fig. 1. The payoff matrix for agent i, $i \in \{1, 2, 3\}$.

The unique Nash equilibrium of this game (when played on any odd cycle) is for all agents to mix between H and T. The agents' payoffs are $\frac{M}{M+1} < 1$.

We analyze this game when all three agents employ a simple regret-minimizing learning dynamic, the replicator dynamics, which can be derived as the continuum limit of the multiplicative-weights learning algorithm. We show that the system does not converge to a set of fixed points but to a globally stable cyclic attractor, the 6-cycle of best responses connecting the 6 pure strategies with social welfare $M + 1$ (see Figure 2), which is significantly higher than the total welfare of < 3 at the unique Nash equilibrium. The analysis follows a delicate line of attack that involves different potential functions on different subsets of the interior of the phase space.

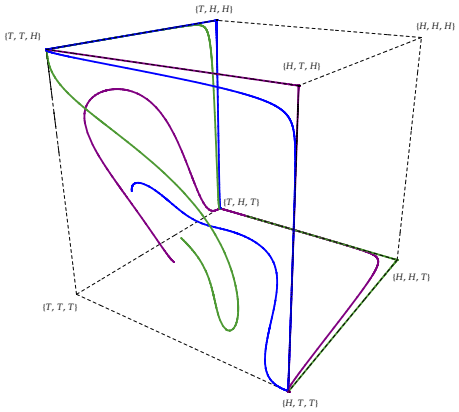

Fig. 2. Convergence to the 6-cycle

[1] Agent numbers are considered to be cyclical, so $0 \equiv 3$.

3. APPLICATIONS TO OLIGOPOLISTIC MARKETS

While the Asymmetric Cyclic Matching Pennies game was designed in a fairly stylized manner, analysis of regret-minimizing algorithms can provide us with rather unexpected insights even in well established settings. In the classic Bertrand model of markets, it is well known that oligopolies with more than two firms exhibit several trivial Nash equilibria. However, in all of these equilibria, the prices are equal to the marginal costs, and all agents make zero profit; this is known as the Bertrand paradox. The same phenomenon occurs for correlated equilibria in the case of a duopoly, where the correlated equilibrium is unique. Recent work [Nadav and Piliouras 2010] studies learning behavior in the Bertrand model, and shows that the zero-profit postulate does not hold for regret-minimizing play, even in the case of two agents. In fact, not only does the market not necessarily converge to zero-profit outcomes, but regret-minimizing agents can actually enjoy significant profits.

So far, we have considered learning behavior under the standard assumption of non-cooperative game theory: agents behave with little regard for the negative externalities they impose on each other. However, in practice, self-interested individuals might explore the possibility of circumventing such negative externalities by forming coalitions.

What sort of coalitions should we expect to arise, and how would they affect the social welfare? Could the outcomes be provably better than the best Nash equilibrium or even the best coalition-less regret-minimizing process?

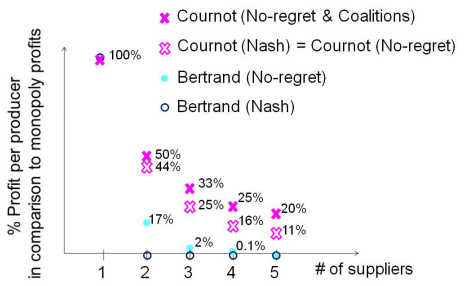

Fig. 3. Regret-minimization and variants lead to profits at least as large as Nash.

Immorlica, Markakis and Piliouras [Immorlica et al. 2010] explore these questions in the classical Cournot model of firm competition. As usual, agents can participate in the market by themselves as singleton coalitions, in which case they can each employ any regret-minimizing strategy of their choice. Furthermore, agents choose strategically how to update the current coalition partition. A new coalition can be created by a merger between two or more coalitions as long as all the participants benefit; an existing coalition can be destroyed by a deviation by a subset of its current players deciding either to form a coalition by themselves or to join another coalition that welcomes them. Finally, the resulting coalitions compete on each round: each coalition acts as a learning-capable entity on behalf of its members and tries to maximize its aggregate utility (which is then split equally among its members) while keeping its regret low.

Immorlica et al. prove tight bounds on the social welfare in this setting which are significantly higher than either that of the unique Nash equilibrium or the best regret-minimizing outcome without coalitions. These bounds are robust across different supply-demand curves and depend only on the size of the market.

4. CONCLUSION

Our community has celebrated many important successes in analyzing Nash equilibria and their properties. Nevertheless, it is important to be aware of the limitations of Nash equilibrium analysis, particularly the limitations of its predictive accuracy. The discussed work highlights this message by showing that in some games, dynamic learning behavior can lead to outcomes that are *much better* than any Nash equilibrium.

It is time to shift our perspective from one that attempts to interpret dynamic behavior mainly in terms of static limit points to one with more refined approaches and techniques. Such a shift, as this exposition underscores, promises exciting new insights as well as novel analytic challenges.

REFERENCES

Immorlica, N., Markakis, E., and Piliouras, G. 2010. Coalition formation and price of anarchy in Cournot oligopolies. In *Workshop on Internet and Network Economics (WINE)*.

Kleinberg, R., Ligett, K., Piliouras, G., and Tardos, E. 2011. Beyond the Nash equilibrium barrier. In *Symposium on Innovations in Computer Science (ICS)*.

Nadav, U. and Piliouras, G. 2010. No-regret learning in oligopolies: Cournot vs Bertrand. In *Symposium on Algorithmic Game Theory (SAGT)*.