Decentralized No-Regret Learning Algorithms for Extensive-Form Correlated Equilibria (Extended Abstract)*

Andrea Celli[1†], **Alberto Marchesi**[1], **Gabriele Farina**[2] and **Nicola Gatti**[1]
[1] Politecnico di Milano, Italy
[2] Carnegie Mellon University, USA
{andrea.celli, alberto.marchesi, nicola.gatti}@polimi.it, gfarina@cs.cmu.edu

Abstract

The existence of uncoupled no-regret learning dynamics converging to correlated equilibria in normal-form games is a celebrated result in the theory of multi-agent systems. Specifically, it has been known for more than 20 years that when all players seek to minimize their *internal* regret in a repeated normal-form game, the empirical frequency of play converges to a normal-form correlated equilibrium. Extensive-form games generalize normal-form games by modeling both sequential and simultaneous moves, as well as imperfect information. Because of the sequential nature and the presence of private information, correlation in extensive-form games possesses significantly different properties than in normal-form games. The extensive-form correlated equilibrium (EFCE) is the natural extensive-form counterpart to the classical notion of correlated equilibrium in normal-form games. Compared to the latter, the constraints that define the set of EFCEs are significantly more complex, as the correlation device (*a.k.a.* mediator) must take into account the evolution of beliefs of each player as they make observations throughout the game. Due to this additional complexity, the existence of uncoupled learning dynamics leading to an EFCE has remained a challenging open research question for a long time. In this article, we settle that question by giving the first uncoupled no-regret dynamics which provably converge to the set of EFCEs in n-player general-sum extensive-form games with perfect recall. We show that each iterate can be computed in time polynomial in the size of the game tree, and that, when all players play repeatedly according to our learning dynamics, the empirical frequency of play after T game repetitions is guaranteed to be a $O(1/\sqrt{T})$-approximate EFCE with high probability, and an EFCE almost surely in the limit.

1 Motivation

This work studies decision-making problems in which *rational* individuals interact with a *centralized planner*. The centralized planner cannot directly tell the individuals what to do, but the goal of the centralized planner is to steer the individuals' behaviors to mutually beneficial outcomes. There are many real-world problems where we observe this type of interaction, and this is increasingly common in the *gig* economy we all live in today. Think, for example, of ride-sharing or food delivery platforms, where drivers provide services to customers, and the whole market is centralized through a single app that every agent connects to.

Because the individual decision makers in the system have free will, the central planner has to take into consideration the fact that all of the individual decision makers will act *selfishly* according to their objectives. Therefore, to get them to behave in a certain way, the central planner must nudge them using the right incentives. This type of soft coordination is already enough to steer the system to social welfare that would be largely impossible in absence of a central planner, so without any form of coordination between the decision makers.

The strategy that the central planner should follow when interacting with decision makers is called a *correlated equilibrium* in the game theory literature. The key feature of a correlated equilibrium is that all the decision makers receive the right incentives to follow the planner's recommendations. This means that no agent would want to do something different from what they are recommended to do by the central planner. The study of this type of equilibrium goes back to the seminal work on correlated equilibrium by Robert Aumann in 1974 [Aumann, 1974], who was later awarded a Nobel prize in economics for his work on game-theoretic cooperation.

Since then, there has been much effort in scaling up the computation of correlated equilibria and designing algorithms guaranteeing some key properties. In particular, a crucial property that algorithms for computing correlated equilibria should satisfy is *decentralization*. This essentially means that the behavior and incentives of each agent should be computed independently from the other agents. The decentralization is fundamental for the computation to scale well, and it allows the agents to converge to an equilibrium point without the need for a central planner. Furthermore, decentralization preserves the agents' privacy during the learning process. Indeed, agents should not need to report their

[*] The complete version of this paper won a best paper award at NeurIPS 2020 [Celli et *al*., 2020]. Some of the results presented here only appear in the full version of the paper [Farina et *al*., 2021].

[†] Contact Author.

true personal preferences back to the central planner or to other decision makers. We also remark that decentralization makes the algorithm robust, since it does not rely on a single point-of-failure. A landmark result by Hart and Mas-Colell in 2000 [Hart and Mas-Colell, 2000] showed that correlated equilibria can be found in such a decentralized way by letting all agents behave independently, according to a simple learning rule.

Unfortunately, the work by Hart and Mas-Colell only applies to one-shot interactions, in which each agent is supposed to interact only once with the system. Agents are not allowed to adjust their behavior based on their observations because they are assumed to act simultaneously, as if it was a rock-paper-scissor kind of interaction. Extending their decentralized approach to the general case where agents act more than once and can adjust their behavior to their observations has been an open question since then. In our work, we close this open problem. This result may foster future applications of game-theoretic solution concepts to real-world decision making, providing ways for individuals to solve hard coordination tasks while pursuing their self-interest.

2 Games, Equilibria, and Learning Dynamics

The *Nash equilibrium* (NE) [Nash, 1950] is the most common notion of rationality in game theory, and its computation in two-player zero-sum games has been the flagship computational challenge in the area at the interplay between computer science and game theory (see, *e.g.*, the landmark results in heads-up no-limit poker, namely [Brown and Sandholm, 2018] and [Moravčík *et al.*, 2017]). The assumption underpinning NE is that the interaction among players is fully *decentralized*. Therefore, an NE is an element of the *uncorrelated* strategy space of the game, that is, a product of independent probability distributions over actions, one per player. A competing notion of rationality is the *correlated equilibrium* (CE) proposed by [Aumann, 1974]. A CE is defined as a probability distribution over joint action profiles—specifying an action for each player—and it is customarily modeled via a trusted external *mediator* that draws an action profile from this distribution, and privately recommends to each player their component. Such probability distribution is a CE if no player has an incentive to choose an action different from the mediator's recommendation, because, assuming that all other players follow their recommended action, the suggested action is the best in expectation.

Many real-world strategic interactions involve more than two players with arbitrary (*i.e.*, general-sum) utilities. In those settings, the CE is an appealing solution concept, as it overcomes several weaknesses of the NE. First, the NE is prone to equilibrium selection issues, raising the question as to how players can select an equilibrium while they are assumed not to be able to communicate with each other. Second, computing an NE is computationally intractable, being PPAD-complete even in two-player games [Chen and Deng, 2006; Daskalakis *et al.*, 2009], whereas a CE can be computed in polynomial time.[1] Third, the social welfare that can be attained via an NE may be arbitrarily lower than what can be achieved via a CE [Koutsoupias and Papadimitriou, 1999; Roughgarden and Tardos, 2002; Celli and Gatti, 2018]. Lastly, in normal-form (that is, simultaneous-move) games, the notion of CE arises from simple *uncoupled* learning dynamics even in general-sum settings with an arbitrary number of players. In words, these learning dynamics are such that each player adjusts their strategy on the basis of their own payoff function, and on other players' strategies, but not on the payoff functions of other players. The existence of uncoupled dynamics enables to overcome the—often unreasonable—assumption that players have perfect knowledge of other players' objectives, while at the same time offering a parallel, scalable avenue for finding equilibria. In contrast, in the case of the NE, uncoupled learning dynamics are only known for the two-player zero-sum setting [Hart and Mas-Colell, 2000; Hart and Mas-Colell, 2003; Cesa-Bianchi and Lugosi, 2006]. The above considerations suggest that CE is oftentimes a better prescriptive solution concept than NE in general-sum and multiplayer settings.

Extensive-form correlated equilibrium (EFCE), introduced by [von Stengel and Forges, 2008], is a natural extension of the correlated equilibrium to the case of extensive-form (that is, sequential) games. Extensive-form games generalize normal-form games by modeling both sequential and simultaneous moves, as well as imperfect information. In an EFCE, the mediator draws, before the beginning of the sequential interaction, a recommended action for each of the possible decision points (also known as, *information sets*) that players may encounter in the game, but these recommendations are not immediately revealed to each player. Instead, the mediator incrementally reveals relevant individual moves as players reach new information sets. At any decision point, the acting player is free to deviate from the recommended action, but doing so comes at the cost of future recommendations, which are no longer issued to that player if they deviate. It is up to the mediator to make sure that the recommended behavior is indeed an equilibrium—that is, that no player would be better off ever deviating from following the mediator's recommendations at each information set. Compared to the constraints that characterize the set of CEs in normal-form games, those that define the set of EFCEs in extensive-form games are significantly more complex. Indeed, the main challenge of the EFCE case is that the mediator must take into account the evolution of beliefs of each player as they make observations throughout the game tree.

In general-sum extensive-form games with an arbitrary number of players (including potentially the *chance player* modeling exogenous stochastic events), the problem of computing a feasible EFCE can be solved in polynomial time in the size of the game tree [Huang and von Stengel, 2008] via a variation of the *Ellipsoid Against Hope* algorithm [Papadimitriou and Roughgarden, 2008; Jiang and Leyton-Brown, 2015]. [Dudík and Gordon, 2009] provide an alternative sampling-based algorithm to compute EFCEs. However,

[1] In normal-form games, a CE can be computed in polynomial time via linear programming. In extensive-form games, the computational complexity of computing a CE depends on the specific notion of correlation that is adopted. The problem can be solved in polynomial time for the notion studied in this article.

their algorithm is centralized and based on MCMC sampling, which limits its applicability on large-scale problems. In practice, these approaches cannot scale beyond toy problems. On the other hand, methods based on uncoupled learning dynamics usually work quite well in large real-world problems, while retaining the appealing properties of uncoupled dynamics that we discussed above.

3 Summary of the Contributions

We focus on the following fundamental research question: *is it possible to devise uncoupled learning dynamics that converge to an EFCE?* We show that the answer is positive.

In the first part of [Celli *et al.*, 2020], we formalize the notion of *trigger regret*, simplifying and extending an idea by [Gordon *et al.*, 2008]. Trigger regret is a notion of regret suitable for extensive-form games that naturally expresses the regret incurred by each player for following the recommendations issued by the EFCE mediator, instead of deviating according to some optimal-in-hindsight strategy. Specifically, trigger regret is a particular instantiation of the framework known as *phi-regret minimization* introduced by [Stoltz and Lugosi, 2007] building on previous work by [Greenwald and Jafari, 2003]. In general, phi-regret minimization operates with a notion of regret defined with respect to a given set of linear transformations on the decision set. In order to define trigger regret, we identify suitable linear transformations that allow us to encode the behavior of trigger agents in the definition of EFCE, which we coin *canonical trigger deviation functions*. Intuitively, canonical trigger deviation functions encode all the possible ways in which a trigger agent may deviate from the recommendations issued by the EFCE mediator, and instead start playing from that point on according to a different strategy than the recommended one. Our core result on trigger regret is the the following: if each player plays according to a no-trigger-regret learning algorithm, then the empirical frequency of play approaches the set of EFCEs.

In the rest of [Celli *et al.*, 2020], we provide an efficient (that is, requiring time polynomial in the size of the game tree at each iteration) algorithm that minimizes trigger regret. The algorithm is based on the general template for constructing phi-regret minimization algorithms given by [Gordon *et al.*, 2008], extending prior work by [Hazan and Kale, 2008]. Before one can use that template, two missing pieces need to be solved:

1. constructing an efficient regret minimizer for the set of all valid canonical trigger deviation functions, and
2. showing that any convex combination of canonical trigger deviation functions admits a fixed point strategy, and that such fixed point can be computed efficiently.

We solve the first point by exploiting the non-trivial combinatorial structures of the set of canonical trigger deviation functions, and the second point by giving an efficient incremental procedure to compute the fixed point strategy in a top-down traversal of the game tree. Our resulting algorithm minimizes trigger regret, guaranteeing $O(\sqrt{T})$ trigger regret with high probability after T iterations and requiring time polynomial in the size of the game tree at each iteration. Thus, when all players play according to the uncoupled learning dynamics defined by our algorithm, the empirical frequency of play after T game repetitions is proven to be a $O(1/\sqrt{T})$-approximate EFCE with high probability, and an EFCE almost surely in the limit. These results generalize the seminal work by [Hart and Mas-Colell, 2000] to the extensive-form game case via a simple and natural framework.

4 Related Works

The study of adaptive procedures leading to a CE dates back to at least the seminal works by [Foster and Vohra, 1997], [Fudenberg and Levine, 1995; Fudenberg and Levine, 1999], and [Hart and Mas-Colell, 2000; Hart and Mas-Colell, 2001]; see also the monograph by [Fudenberg and Levine, 1998]. In particular, the work by [Hart and Mas-Colell, 2000] proves that simple dynamics based on the notion of *internal regret* converge to a CE in normal-form games. The strategy that the authors introduce—the so-called *regret matching*—is conceptually simple, and guarantees that if all players follow this strategy, then the empirical frequency of play converges to the set of CEs (see also [Cahn, 2004]). Other works describe extensions to the models studied in the aforementioned papers. For example, [Stoltz and Lugosi, 2007] describe an adaptive procedure converging to a CE in games with an infinite, but compact, set of actions, while [Kakade *et al.*, 2003] consider efficient algorithms for computing correlated equilibria in graphical games.

In more recent years, a growing effort has been devoted to understanding the relationships between no-regret learning dynamics and equilibria in extensive-form games. These games pose additional challenges when compared to normal-form games, due to their sequential nature and the presence of imperfect information. While in two-player zero-sum extensive-form games it is widely known that no-regret learning dynamics converge to an NE—with the *counterfactual regret minimization* (CFR) algorithm and its variations being the state of the art for equilibrium finding in such games [Zinkevich *et al.*, 2008; Tammelin, 2014; Tammelin *et al.*, 2015; Lanctot *et al.*, 2009; Brown and Sandholm, 2019]—the general case is less understood. [Celli *et al.*, 2019] provide some variations of the classical CFR algorithm for n-player general-sum extensive-form games, showing that they provably converge to a *normal-form coarse correlated equilibrium*, which is based on a form of correlation that is less appealing than that of EFCE in sequential games.

Finally, we mention relevant literature subsequent to the conference version of this article. In a recent paper, [Morrill *et al.*, 2020] conduct a study of different forms of correlation in extensive-form games, defining a taxonomy of solution concepts. Each of their solution concepts is attained by a particular set of no-regret learning dynamics, which is obtained by instantiating the phi-regret minimization framework [Greenwald and Jafari, 2003; Stoltz and Lugosi, 2007; Gordon *et al.*, 2008] with a suitably-defined deviation function. As part of their analysis, [Morrill *et al.*, 2020] investigate some properties of the well-established CFR regret minimization algorithm [Zinkevich *et al.*, 2008] applied to n-player general-sum extensive-form games, establishing that it is hindsight-rational with respect to a specific set of deviation functions, which the authors coin *blind counterfactual deviations*. Moreover, in a very recent working paper, [Morrill *et al.*, 2021] extend their prior work [Morrill *et al.*, 2020] by identifying a general class of deviations—called *behavioral deviations*—that induce equilibria that can be found through uncoupled no-regret learning dynamics.

References

[Aumann, 1974] Robert J Aumann. Subjectivity and correlation in randomized strategies. *Journal of mathematical Economics*, 1(1):67–96, 1974.

[Brown and Sandholm, 2018] Noam Brown and Tuomas Sandholm. Superhuman AI for heads-up no-limit poker: Libratus beats top professionals. *Science*, 359(6374):418–424, 2018.

[Brown and Sandholm, 2019] Noam Brown and Tuomas Sandholm. Solving imperfect-information games via discounted regret minimization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 1829–1836, 2019.

[Cahn, 2004] Amotz Cahn. General procedures leading to correlated equilibria. *International Journal of Game Theory*, 33(1):21–40, 2004.

[Celli and Gatti, 2018] A. Celli and N. Gatti. Computational results for extensive-form adversarial team games. In *AAAI Conference on Artificial Intelligence (AAAI)*, 2018.

[Celli *et al.*, 2019] Andrea Celli, Alberto Marchesi, Tommaso Bianchi, and Nicola Gatti. Learning to correlate in multi-player general-sum sequential games. In *Advances in Neural Information Processing Systems*, pages 13055–13065, 2019.

[Celli *et al.*, 2020] Andrea Celli, Alberto Marchesi, Gabriele Farina, and Nicola Gatti. No-regret learning dynamics for extensive-form correlated equilibrium. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2020.

[Cesa-Bianchi and Lugosi, 2006] Nicolo Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge University Press, 2006.

[Chen and Deng, 2006] Xi Chen and Xiaotie Deng. Settling the complexity of two-player nash equilibrium. In *2006 47th Annual IEEE Symposium on Foundations of Computer Science*, pages 261–272. IEEE, 2006.

[Daskalakis *et al.*, 2009] Constantinos Daskalakis, Paul W Goldberg, and Christos H Papadimitriou. The complexity of computing a nash equilibrium. *SIAM Journal on Computing*, 39(1):195–259, 2009.

[Dudík and Gordon, 2009] Miroslav Dudík and Geoffrey J Gordon. A sampling-based approach to computing equilibria in succinct extensive-form games. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, pages 151–160, 2009.

[Farina *et al.*, 2021] Gabriele Farina, Andrea Celli, Alberto Marchesi, and Nicola Gatti. Simple uncoupled no-regret learning dynamics for extensive-form correlated equilibrium. *arXiv preprint arXiv:2104.01520*, 2021.

[Foster and Vohra, 1997] Dean P Foster and Rakesh V Vohra. Calibrated learning and correlated equilibrium. *Games and Economic Behavior*, 21(1-2):40, 1997.

[Fudenberg and Levine, 1995] Drew Fudenberg and David K Levine. Consistency and cautious fictitious play. *Journal of Economic Dynamics and Control*, 19(5-7):1065–1089, 1995.

[Fudenberg and Levine, 1998] Drew Fudenberg and David K Levine. *The theory of learning in games*, volume 2. MIT press, 1998.

[Fudenberg and Levine, 1999] Drew Fudenberg and David K Levine. Conditional universal consistency. *Games and Economic Behavior*, 29(1-2):104–130, 1999.

[Gordon *et al.*, 2008] Geoffrey J Gordon, Amy Greenwald, and Casey Marks. No-regret learning in convex games. In *International Conference on Machine learning*, pages 360–367, 2008.

[Greenwald and Jafari, 2003] Amy Greenwald and Amir Jafari. A general class of no-regret learning algorithms and game-theoretic equilibria. In *Learning theory and kernel machines*, pages 2–12. Springer, 2003.

[Hart and Mas-Colell, 2000] Sergiu Hart and Andreu Mas-Colell. A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68(5):1127–1150, 2000.

[Hart and Mas-Colell, 2001] Sergiu Hart and Andreu Mas-Colell. A general class of adaptive strategies. *Journal of Economic Theory*, 98(1):26–54, 2001.

[Hart and Mas-Colell, 2003] Sergiu Hart and Andreu Mas-Colell. Uncoupled dynamics do not lead to nash equilibrium. *The American Economic Review*, 93(5):1830–1836, 2003.

[Hazan and Kale, 2008] Elad Hazan and Satyen Kale. Computational equivalence of fixed points and no regret algorithms, and convergence to equilibria. In *Advances in Neural Information Processing Systems*, volume 20, 2008.

[Huang and von Stengel, 2008] Wan Huang and Bernhard von Stengel. Computing an extensive-form correlated equilibrium in polynomial time. In *International Workshop on Internet and Network Economics*, pages 506–513. Springer, 2008.

[Jiang and Leyton-Brown, 2015] Albert Xin Jiang and Kevin Leyton-Brown. Polynomial-time computation of exact correlated equilibrium in compact games. *Games and Economic Behavior*, 91:347–359, 2015.

[Kakade *et al.*, 2003] Sham Kakade, Michael Kearns, John Langford, and Luis Ortiz. Correlated equilibria in graphical games. In *Proceedings of the 4th ACM Conference on Electronic Commerce*, pages 42–47, 2003.

[Koutsoupias and Papadimitriou, 1999] Elias Koutsoupias and Christos Papadimitriou. Worst-case equilibria. In *Annual Symposium on Theoretical Aspects of Computer Science*, pages 404–413. Springer, 1999.

[Lanctot *et al.*, 2009] Marc Lanctot, Kevin Waugh, Martin Zinkevich, and Michael H Bowling. Monte carlo sampling for regret minimization in extensive games. In *Advances in Neural Information Processing Systems*, pages 1078–1086, 2009.

[Moravčík *et al.*, 2017] Matej Moravčík, Martin Schmid, Neil Burch, Viliam Lisỳ, Dustin Morrill, Nolan Bard, Trevor Davis, Kevin Waugh, Michael Johanson, and Michael Bowling. Deepstack: Expert-level artificial intelligence in heads-up no-limit poker. *Science*, 356(6337):508–513, 2017.

[Morrill *et al.*, 2020] Dustin Morrill, Ryan D'Orazio, Reca Sarfati, Marc Lanctot, James Wright, Amy Greenwald, and Michael Bowling. Hindsight and sequential rationality of correlated play. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI-20)*, 2020.

[Morrill *et al.*, 2021] Dustin Morrill, Ryan D'Orazio, Marc Lanctot, James R Wright, Michael Bowling, and Amy Greenwald. Efficient deviation types and learning for hindsight rationality in extensive-form games. *arXiv preprint arXiv:2102.06973*, 2021.

[Nash, 1950] John F Nash. Equilibrium points in n-person games. *Proceedings of the national academy of sciences*, 36(1):48–49, 1950.

[Papadimitriou and Roughgarden, 2008] Christos H Papadimitriou and Tim Roughgarden. Computing correlated equilibria in multi-player games. *Journal of the ACM*, 55(3):14, 2008.

[Roughgarden and Tardos, 2002] Tim Roughgarden and Éva Tardos. How bad is selfish routing? *Journal of the ACM*, 49(2):236–259, 2002.

[Stoltz and Lugosi, 2007] Gilles Stoltz and Gábor Lugosi. Learning correlated equilibria in games with compact sets of strategies. *Games and Economic Behavior*, 59(1):187–208, 2007.

[Tammelin *et al.*, 2015] Oskari Tammelin, Neil Burch, Michael Johanson, and Michael Bowling. Solving heads-up limit texas hold'em. In *International Joint Conferences on Artificial Intelligence*, pages 645–652, 2015.

[Tammelin, 2014] Oskari Tammelin. Solving large imperfect information games using cfr+. *arXiv preprint arXiv:1407.5042*, 2014.

[von Stengel and Forges, 2008] Bernhard von Stengel and Françoise Forges. Extensive-form correlated equilibrium: Definition and computational complexity. *Mathematics of Operations Research*, 33(4):1002–1022, 2008.

[Zinkevich *et al.*, 2008] Martin Zinkevich, Michael Johanson, Michael Bowling, and Carmelo Piccione. Regret minimization in games with incomplete information. In *Advances in Neural Information Processing Systems*, pages 1729–1736, 2008.