MIT 6.S890 — Topics in Multiagent Learning

Lecture 6
Learning in games: Algorithms (Part II)

Instructor: Prof. Gabriele Farina

1 Predictivity, optimism, and acceleration

In recent years, there has been a lot of interest in the idea of *optimism* in learning algorithms. The fundamental idea behind optimism is the following:

> *When all players learn at the same time, the environment is nonstationary but not necessarily adversarial. Can one then take advantage of this to design learning algorithms with better regret guarantees and convergence properties?*

1.1 Predictive algorithms

The idea of predictive is to *anticipate* the next utility gradient $g^{(t+1)}$ by having a *prediction* $m^{(t+1)}$. Based on the prediction, the three following predictive algorithms can be defined.

	Non-predictive version	Predictive version
FTRL	$x^{(t+1)} := \arg\max_{x\in\mathcal{X}}\left\{\left\langle \sum_{\tau=1}^{t} g^{(\tau)}, x\right\rangle - \frac{1}{\eta}\psi(x)\right\}$	$x^{(t+1)} := \arg\max_{x\in\mathcal{X}}\left\{\left\langle m^{(t+1)} + \sum_{\tau=1}^{t} g^{(\tau)}, x\right\rangle - \frac{1}{\eta}\psi(x)\right\}$
OMD	$x^{(t+1)} := \arg\max_{x\in\mathcal{X}}\left\{\langle g^{(t)}, x\rangle - \frac{1}{\eta}\mathrm{D}_\psi(x \,\|\, x^{(t)})\right\}$	• **Non-reflected version:** $z^{(t+1)} := \arg\max_{z\in\mathcal{X}}\left\{\langle g^{(t)}, z\rangle - \frac{1}{\eta}\mathrm{D}_\psi(z \,\|\, z^{(t)})\right\}$ $x^{(t+1)} := \arg\max_{x\in\mathcal{X}}\left\{\langle m^{(t+1)}, x\rangle - \frac{1}{\eta}\mathrm{D}_\psi(x \,\|\, z^{(t+1)})\right\}$ • **Reflected version:** $x^{(t+1)} := \arg\max_{x\in\mathcal{X}}\left\{\langle g^{(t)} + m^{(t+1)} - m^{(t)}, x\rangle - \frac{1}{\eta}\mathrm{D}_\psi(x \,\|\, x^{(t)})\right\}$

While the three predictive algorithms are in general different, they coincide in the special case of *Legendre regularizers* (see also Remark 2.1 in Lecture 5).

Remark 1.1. For Legendre regularizers (*i.e.*, when ψ's gradients go to infinity at the boundary of $\mathcal{X}$), the three predictive algorithms (FTRL, non-reflected OMD, reflected OMD) coincide.

It is also worth noting that the predictive versions of the algorithms subsume the non-predictive versions as a special case, as we point out in the next remark.

⋆These notes are class material that has not undergone formal peer review. The TAs and I are grateful for any reports of typos.

Remark 1.2. The standard (non-predictive) FTRL and OMD algorithms correspond to the case where the prediction is set to zero, *i.e.*, $m^{(t)} = 0$ at all times t.

1.2 Optimism

The idea of optimism is to use predictivity with the specific guess $m^{(t+1)} = g^{(t)}$ at all times t. This corresponds to predicting that the feedback is slow-changing.

Example 1.1 (Optimistic online gradient ascent). The non-reflected OMD algorithm instantiated with squared Euclidean norm $\psi(x) = \frac{1}{2}\|x\|_2^2$ gives rise to the (non-reflected) *optimistic online gradient ascent* algorithm, whose update rule is

$$z^{(t+1)} := \prod_{\mathcal{X}}\big(z^{(t)} + \eta g^{(t)}\big), \qquad x^{(t+1)} := \prod_{\mathcal{X}}\big(z^{(t+1)} + \eta g^{(t)}\big). \tag{1}$$

Example 1.2 (Optimistic MWU). For the MWU algorithm, the optimistic version of FTRL, non-reflected OMD, and reflected OMD all coincide, and give rise to the following update rule:

$$x^{(t+1)} \propto \exp\big(\eta r^{(t)} + \eta\big(r^{(t)} - r^{(t-1)}\big)\big).$$

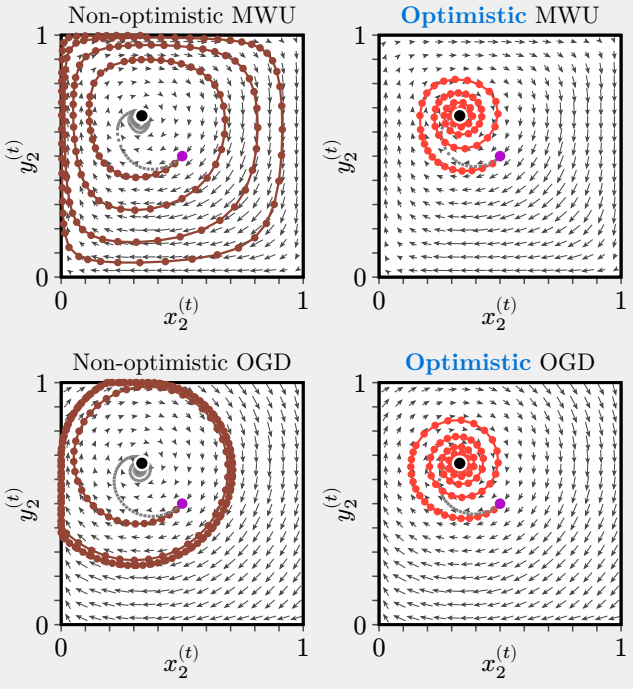
In two-player games, optimism serves as a form of *negative* momentum that pushes the iterates towards the equilibrium. We illustrate this in the following example. We will give a quantitative analysis of the effect of optimism in the convergence of learning algorithms in Section 2.

Example 1.3. The plots on the right show the dynamics of the optimistic and non-optimistic versions of MWU and OGD in the small two-player zero-sum game we used in Example 2.2 of Lecture 5, whose utility matrix is

$$\mathrm{U}_1 := \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}.$$

The multiplicative weights update algorithm was set up with constant learning rate $\eta = 0.25$, while the online gradient descent algorithm was set up with learning rate $\eta = 0.1$. The optimistic version of online projected gradient descent (OGD) was non-reflected. The purple dot indicates the starting strategy. The gray dotted line tracks the profile of *average* strategies.

As mentioned, the optimistic dynamics exhibit a "push" towards equilibrium, due to the negative momentum effect, which results in convergence towards the unique Nash equilibrium $x^* = (2/3, 1/3)$, $y^* = (1/3, 2/3)$.



1.3 Predictive regret bounds (RVU)

Intuitively, one would expect that predictions *help* in reducing the regret of the learning algorithm. At one extreme, one would presumably hope that if the prediction is perfect, then the regret would be very small. This is indeed the case, as shown by [Syr+15].

Theorem 1.1 (RVU bound, [Syr+15]). Predictive FTRL and Predictive OMD satisfy the following regret bound, which is often called *RVU bound* (regret bounded by variation in utilities):

$$\mathrm{Reg}^{(T)} \leq \max_{\hat{x} \in \mathcal{X}} \frac{\psi(\hat{x}) - \psi\big(x^{(1)}\big)}{\eta} + \eta \sum_{t=1}^{T} \left\| g^{(t)} - m^{(t)} \right\|_*^2 - \frac{1}{8\eta} \sum_{t=2}^{T} \left\| x^{(t)} - x^{(t-1)} \right\|^2,$$

where $\|\cdot\|_*$ is the dual norm of $\|\cdot\|$.

Remark 1.3. A consequence of the previous regret bound is the fact that—assuming $m^{(t)} = g^{(t)}$ is *omniscient*—the regret of the learning algorithm *does not grow with time*.

1.4 Accelerated convergence to Nash equilibria in two-player zero-sum games

As noted by [Syr+15], the RVU bound implies accelerated convergence to Nash equilibria in two-player zero-sum games. The proof is quite elementary, and we present it next.

Theorem 1.2 (Accelerated convergence to Nash equilibria in two-player zero-sum games, [Syr+15]). Consider any two-player zero-sum game; let $\mathrm{U}_1 \in \mathbb{R}^{m \times n}$ be the utility matrix for Player 1. If the players employ regret minimizers that guarantee RVU regret bounds of the form

$$\mathrm{Reg}_1^{(T)} \leq \frac{\Omega_1}{\eta} + \eta \sum_{t=1}^{T} \left\| \mathrm{U}_1 \big(y^{(t)} - y^{(t-1)}\big) \right\|_*^2 - \frac{1}{8\eta} \sum_{t=1}^{T} \left\| x^{(t)} - x^{(t-1)} \right\|^2$$

$$\mathrm{Reg}_2^{(T)} \leq \frac{\Omega_2}{\eta} + \eta \sum_{t=1}^{T} \left\| \mathrm{U}_1{}^{\top} \big(x^{(t)} - x^{(t-1)}\big) \right\|_*^2 - \frac{1}{8\eta} \sum_{t=1}^{T} \left\| y^{(t)} - y^{(t-1)} \right\|^2,$$

and $\eta \leq 1/\Big(4\|\mathrm{U}_1\|_{\mathrm{op}}\Big)$, where $\|\mathrm{U}_1\|_{\mathrm{op}} := \max_{z \in \mathbb{R}^n} \|\mathrm{U}_1 z\|_* / \|z\|$ is the operator norm of U_1, then, at any time T, the sum of the regrets of the players satisfies the bound

$$\mathrm{Reg}_1^{(T)} + \mathrm{Reg}_2^{(T)} \leq \frac{\Omega_1 + \Omega_2}{\eta}$$

which is *constant* with respect to time. This immediately implies convergence to the set of Nash equilibria in two-player zero-sum games at the rate of $O_T(1/T)$.

Proof. The statement follows from summing up the RVU bounds, and observing that the middle terms cancel out with the right-most terms. More precisely, we have

$$\mathrm{Reg}_1^{(T)} \leq \frac{\Omega_1}{\eta} + \eta \|\mathrm{U}_1\|_{\mathrm{op}} \sum_{t=1}^{T} \left\| y^{(t)} - y^{(t-1)} \right\|_*^2 - \frac{1}{8\eta} \sum_{t=1}^{T} \left\| x^{(t)} - x^{(t-1)} \right\|^2$$

$$\mathrm{Reg}_2^{(T)} \leq \frac{\Omega_2}{\eta} + \eta \|\mathrm{U}_1\|_{\mathrm{op}} \sum_{t=1}^{T} \left\| x^{(t)} - x^{(t-1)} \right\|_*^2 - \frac{1}{8\eta} \sum_{t=1}^{T} \left\| y^{(t)} - y^{(t-1)} \right\|^2$$

Summing and using the fact that $\eta \leq 1/4\|\mathrm{U}_1\|_{\mathrm{op}}$ by assumption, we obtain the statement. □

1.5 Accelerated convergence to coarse correlated equilibria in general games

Rates of $\tilde{O}(1/T)$ for coarse correlated and correlated equilibria (CCE) via learning in normal-form games (and beyond) are also known for the multiplayer case, but they are significantly harder to prove. One of the main obstacles is due to the fact that convergence to CCE is driven by the *maximum* of the regrets of the players, and not the sum as in two-player zero-sum Nash equilibria.

We mention some of the results in this direction.

- Syrgkanis, V., Agarwal, A., Luo, H., & Schapire, R. E. [Syr+15] showed $O\Big(n \log|A| \; T^{-\frac{3}{4}}\Big)$ for OMWU using RVU bounds.
- This result was later improved by Chen, X., & Peng, B. [CP20] to $O\Big(n \log^{\frac{5}{6}}|A| \; T^{-\frac{5}{6}}\Big)$ for *two-player* general-sum games only.
- Daskalakis, C., Fishelson, M., & Golowich, N. [DFG21] showed $O\Big(n \log|A| \; \frac{\log^4 T}{T}\Big)$ convergence for OMWU using a very complicated analysis based on the idea of high-order stability.
- Farina, G., Anagnostides, I., Luo, H., Lee, C.-W., Kroer, C., & Sandholm, T. [Far+22] showed $O\Big(n \; |A| \; \frac{\log T}{T}\Big)$ convergence rates using RVU bounds paired with a special regularizer.

2 Convergence in iterates

The convergence results we have seen so far pertain to the *average* strategies (either individual, or the average of the product) produced by learning dynamics. One might then wonder what is known about the *iterate* convergence to equilibrium. Complexity-theoretic considerations regarding the hardness of approximating Nash equilibria preclude this phenomenon beyond two-player zero-sum games. As we now argue, in two-player zero-sum games the phenomenon is indeed possible.

■ **Best-iterate convergence.** Anagnostides, I., Panageas, I., Farina, G., & Sandholm, T. [Ana+22] showed that when both players use optimistic gradient ascent, the *best* iterate converges to the Nash equilibrium in two-player zero-sum games at the rate of $O_T\Big(1/\sqrt{T}\Big)$. At a high level, the proof of this result is in two steps. First, the authors show that the sum of the squared distances between consecutive iterates is bounded by a constant. This implies that at least one iterate is close to the previous one. Second, they show that small simultaneous movements imply proximity to a Nash equilibrium. Both of these steps require only elementary calculations; feel free to try to reproduce the result yourself or check the details in the original paper.

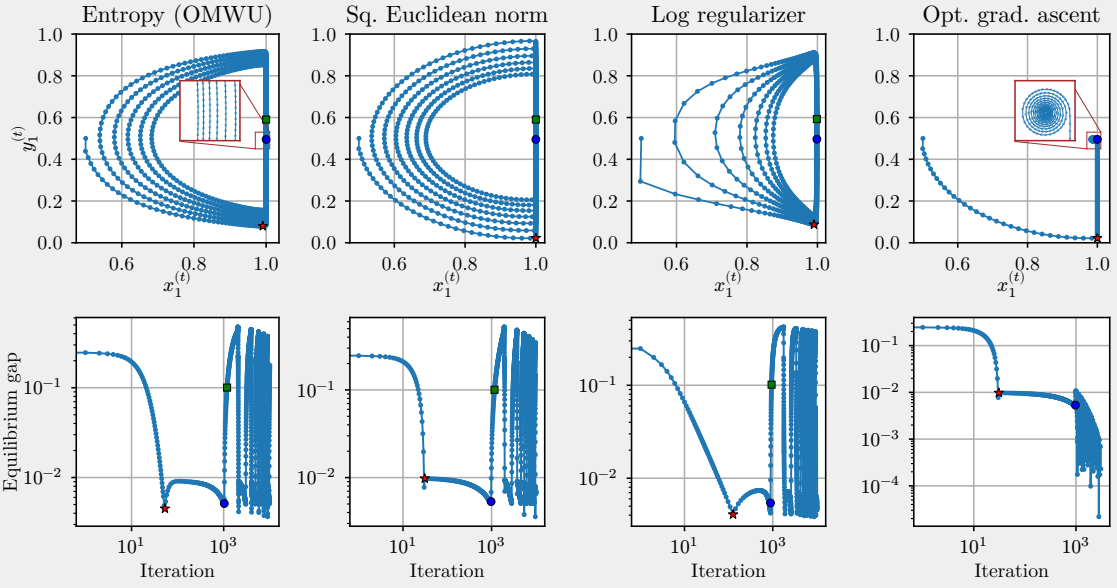
■ **Last-iterate convergence.** Cai, Y., Oikonomou, A., & Zheng, W. [COZ22] improved the best-iterate result mentioned above by showing *last* iterate converges to the Nash equilibrium at the rate of $O_T\Big(1/\sqrt{T}\Big)$ for optimistic OGD. Their analysis is significantly more involved, and revolves around studying a Lyapunov potential function that was discovered via semidefinite programming.

Both of the results mentioned above pertain to optimistic OGD. Arguably, it was believed for a while in the community that good last-iterate convergence of OMWU were in the air, just "one good trick" away. After all, OMWU had always spoiled us with its good properties. Furthermore, the paper by Hsieh, Y.-G., Antonakopoulos, K., & Mertikopoulos, P. [HAM21] showed *asymptotic* (*i.e.*, in the limit, but without any concrete rates) convergence of optimistic MWU to the set of equilibria in two-player zero-sum games. So, it seemed pretty likely that good, concrete rates of convergence could be established beyon optimistic gradient ascent. However, in a recent twist, it was shown that the sitution is less rosy than expected. We illustrate this with an example.

Example 2.1 (Poor last-iterate convergence of FTRL, [Cai+24]). Consider the two-player zero-sum game with utility matrix for Player 1 given by

$$\mathrm{U}_1(\delta) \coloneqq \begin{pmatrix} \frac{1}{2}+\delta & \frac{1}{2} \\ 0 & 1 \end{pmatrix}.$$

The game admits the unique Nash equilibrium (x^*, y^*) where $x^* = \left(\frac{1}{1+\delta}, \frac{\delta}{1+\delta}\right), y^* = \left(\frac{1}{2(1+\delta)}, \frac{1+2\delta}{2(1+\delta)}\right)$. In particular, when δ is small, the equilibrium strategy for Player 1 is approximately $x^* = (1-\delta, \delta)$ and thus very close to the boundary of the strategy polytope of the player. This proximity to the boundary affects the performance of all known instantiations of the the optimistic FTRL algorithm. To see this numerically, the next four plots show the evolution of three optimistic FTRL variants (entropic, Euclidean, and logarithmic) and the optimistic gradient ascent algorithm, in the game defined by $\delta = 10^{-2}$.



The dynamics for the the first two algorithms get *extremely* close to the boundary—for example, when using OMWU, iterates reached strategies with $1 - e^{-50} < x_1 < 1$.

By studying the dynamics produced by instantiations of the FTRL algorithm in the previous game, the following result can be established.

Theorem 2.1 (Informal, [Cai+24]). Under standard assumptions about the regularizer, there is no function f such that optimistic FTRL produces a last-iterate convergence rate of $f(|A_1|, |A_2|, T) \to 0$ when all payoffs of the game are in $[0, 1]$, and $|A_1|$ and $|A_2|$ are the number of actions of the players. In other words, the last-iterate convergence rate must depend on *some form of condition number* of the game.

Bibliography

[Syr+15] V. Syrgkanis, A. Agarwal, H. Luo, and R. E. Schapire, "Fast convergence of regularized learning in games," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 28, 2015.

[CP20] X. Chen and B. Peng, "Hedging in games: Faster convergence of external and swap regrets," *Advances in Neural Information Processing Systems*, vol. 33, pp. 18990–18999, 2020.

[DFG21] C. Daskalakis, M. Fishelson, and N. Golowich, "Near-optimal no-regret learning in general games," *Advances in Neural Information Processing Systems*, vol. 34, pp. 27604–27616, 2021.

[Far+22] G. Farina, I. Anagnostides, H. Luo, C.-W. Lee, C. Kroer, and T. Sandholm, "Near-optimal no-regret learning dynamics for general convex games," *Advances in Neural Information Processing Systems*, vol. 35, pp. 39076–39089, 2022.

[Ana+22] I. Anagnostides, I. Panageas, G. Farina, and T. Sandholm, "On Last-Iterate Convergence Beyond Zero-Sum Games," in *International Conference on Machine Learning*, 2022.

[COZ22] Y. Cai, A. Oikonomou, and W. Zheng, "Finite-Time Last-Iterate Convergence for Learning in Multi-Player Games," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2022. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2022/file/db2d2001f63e83214b08948b459f69f0-Paper-Conference.pdf

[HAM21] Y.-G. Hsieh, K. Antonakopoulos, and P. Mertikopoulos, "Adaptive learning in continuous games: Optimal regret bounds and convergence to nash equilibrium," in *Conference on Learning Theory*, 2021, pp. 2388–2422.

[Cai+24] Y. Cai *et al.*, "Fast Last-Iterate Convergence of Learning in Games Requires Forgetful Algorithms," *arXiv*, Jun. 2024, doi: 10.48550/arXiv.2406.10631.

Changelog

- Sep 24: typos fixed.
- Sep 24: added Example 1.3.
- Sep 25: expanded discussion on constant regret and best-iterate convergence.
- Sep 27: fixed typo.

6.S890 Topics in Multiagent Learning

Fall 2024