Lecture 11: Optimistic Online Learning and Social Welfare of No-Regret Dynamics

Lecturer: Anas Barakat

October 21, 2025

Abstract

In the first part, we show how optimism can be used in FTRL to to establish improved regret bounds in the game setting, in contrast with the adversarial setting. The analysis relies on bounding the regret by variation in utilities and exploiting the smooth variation of payoff vectors when using a strongly-convex regularizer in Optimistic FTRL with suitable small enough step sizes. In the second part, we discuss social welfare of no-regret dynamics via the class of smooth games, establishing a lower bound of the time-average welfare of the sequence of play with respect to the optimal welfare of the static game.

Disclaimer: These lecture notes are preliminary notes that will evolve and will be updated over time, this is the first iteration of the course. The treatment is not comprehensive but focuses on some of the main ideas with accompanying proofs. Research on the topic is also still very active.

1 Preamble: No-Regret Learning in Games

Game setting. Consider a game $\Gamma = (\mathcal{I}, \{\mathcal{X}_i\}_{i\in\mathcal{I}}, \{u_i\}_{i\in\mathcal{I}})$ where:

- $\mathcal{I}$ is a finite set of $|\mathcal{I}| = N$ players,
- $\mathcal{X}_i = \Delta(\mathcal{A}_i)$ is the decision set of each player $i \in \mathcal{I}$ which is the simplex over the finite set of actions $\mathcal{A}_i$,
- $u_i : \prod_{j=1}^N \mathcal{X}_j \to \mathbb{R}$ is the utility function of player $i \in \mathcal{I}$ which is multilinear (i.e. linear in each one of the variables when other variables are fixed).

Remark 1. The above setting can also capture extensive-form games and more generally convex games using compact convex sets $\mathcal{X}_i$.

Repeated game setting and regret. We consider a repeated game setting where the game Γ is played repeatedly for T rounds. At each time step, each agent $i \in \mathcal{I}$ selects a strategy $x_i^t \in \mathcal{X}_i$ and observes a payoff vector u_i^t (e.g. $u_i^t := (u_i(a_i, x_{-i}^t))_{a_i\in\mathcal{A}_i}$ in normal-form games). Recall then that the (external)-regret of each player $i \in \mathcal{I}$ is then defined as follows:

$$R_i^T := \max_{x_i^\star\in\mathcal{X}_i} \sum_{t=1}^T \langle x_i^\star - x_i^t, u_i^t\rangle = \underbrace{\max_{x_i^\star\in\mathcal{X}_i} \left\{\sum_{t=1}^T \langle x_i^\star, u_i^t\rangle\right\}}_{\text{comparator}} - \underbrace{\sum_{t=1}^T \langle x_i^t, u_i^t\rangle}_{\text{cumulative payoffs}} . \tag{1}$$

Remark 2. Recall that regret can be negative in general.

So far, in previous lectures we have discussed convergence to Nash equilibria in potential games and two-player zero-sum games, as well as to (coarse) correlated equilibria in multi-player general-sum games with a $\tilde{\mathcal{O}}(1/\sqrt{T})$ rate of convergence. This convergence rate is based on the worst-case (under adversarial losses in regret minimization) guarantee of $\tilde{\mathcal{O}}(\sqrt{T})$ for regret upper-bounds using no-regret algorithms.

However, in the game setting when all players use a no-regret algorithm, the payoff vectors do not evolve in a completely adversarial/unpredictable manner. It is hence natural to ask whether we can improve our analysis in the present game setting.

In this lecture, we ask whether we can establish faster convergence rates than $\tilde{\mathcal{O}}(1/\sqrt{T})$.

2 Optimistic Online Learning

2.1 Can we achieve faster rates using Multiplicative Weights Updates?

Theorem 3 (Chen and Peng (2020)). *For any learning rate, when both players in a two-player game use MWU, at least one player will suffer $\Omega(\sqrt{T})$ regret.*

2.2 Optimistic Follow-The-Regularized-Leader

The main idea is to use a prediction of the next utility vector in the update rule (Chiang et al., 2012; Rakhlin and Sridharan, 2013).

$$x_i^{t+1} = \operatorname*{argmax}_{x_i^* \in \mathcal{X}_i} \left\{ \left\langle x_i^*, \sum_{\tau=1}^{t} u_i^\tau + m_i^{t+1} \right\rangle - \frac{1}{\eta} \mathcal{R}(x_i^*) \right\}, \qquad \text{(Optimistic FTRL)}$$

where $\mathcal{R}$ is a regularizer, η is a positive step size and m_i^{t+1} is a predictive sequence, typically chosen as $m_i^{t+1} = u_i^t$.[1]

Remark 4. Setting $m_i^t = 0$ recovers the classical (non-optimistic) FTRL algorithm.

2.3 Regret Bounded by Variation in Utilities

Typical regret bounds (e.g. for FTRL) showed in the previous lectures are as follows:

$$R_i^T \leq \frac{\alpha}{\eta} + \eta \sum_{t=1}^{T} \|u_i^t\|_\star^2, \tag{2}$$

where $(\|\cdot\|, \|\cdot\|_\star)$ are dual norms.

The following result provides a refined regret bound where the first error term captures the prediction error induced by using *optimistic* FTRL.

Theorem 5 (RVU bound, Syrgkanis et al. (2015)). *The sequence generated by Optimistic FTRL satisfies the following bound for any $T \geq 1$:*

$$R_i^T \leq \frac{\alpha}{\eta} + \beta\eta \sum_{t=1}^{T} \|u_i^t - m_i^t\|_*^2 - \frac{\gamma}{\eta} \sum_{t=1}^{T} \|x_i^t - x_i^{t-1}\|^2, \tag{3}$$

where $\alpha = R$ with $R := \max_{i \in \mathcal{I}} (\sup_{x_i \in \Delta(\mathcal{A}_i)} \mathcal{R}(x_i) - \inf_{x_i \in \Delta(\mathcal{A}_i)} \mathcal{R}(x_i))$, $\beta = 1$ and $\gamma = 1/4$.

Remark 6. Optimistic Mirror Descent (which we do not introduce in this lecture) also satisfies a similar RVU bound with different constants (α, β, γ), see Syrgkanis et al. (2015).

Lemma 7. *For any $i \in \mathcal{I}$ and any $t \geq 1$,*

$$\|u_i^t - u_i^{t-1}\|_*^2 \leq (N-1) \sum_{j \neq i} \|x_j^t - x_j^{t-1}\|^2. \tag{4}$$

Proof. We assume that $u_i(x) \leq 1$ for any $x \in \mathcal{X}$ (without loss of generality). Using the definition of the payoff vectors and the boundedness of the payoffs, we have:

$$\left\|u_i^t - u_i^{t-1}\right\|_* \leq \sum_{a_{-i} \in \mathcal{A}_{-i}} \left| \prod_{j \neq i} x_{j,a_j}^t - \prod_{j \neq i} x_{j,a_j}^{t-1} \right| \leq \sum_{j \neq i} \left\|x_j^t - x_j^{t-1}\right\|,$$

[1] More sophisticated predictive sequences (such as averages of past utility vectors over a given past window of time) can also be chosen.

where the latter inequality follows from bounding the total variation distance of two product distributions by the sum of the total variations of each marginal distribution. Using Jensen's inequality, it follows that:

$$\|u_i^t - u_i^{t-1}\|_*^2 \leq \left(\sum_{j\neq i}\|x_j^t - x_j^{t-1}\|\right)^2 \leq (N-1)\sum_{j\neq i}\|x_j^t - x_j^{t-1}\|^2,$$

which concludes the proof. □

Lemma 8 (Stability)**.** *The following statements hold.*

(i) If player $i \in \mathcal{I}$ implements optimistic FTRL, then $\|x_i^t - x_i^{t-1}\| = \mathcal{O}(\eta)$.

(ii) If all players implement optimistic FTRL, then $\|u_i^t - u_i^{t-1}\|_* = \mathcal{O}(\eta)$.

Proof. The second item follows from the first one by using Lemma 7. We prove the first item.[2]

Denote by $\mathcal{R}^\star$ the Fenchel conjugate function defined by:

$$\mathcal{R}^\star(y) = \sup_{x\in\mathcal{X}}\{\langle y, x\rangle - \mathcal{R}(x)\}. \tag{5}$$

Lemma 9. *The function $\mathcal{R} : \mathcal{X} \to \mathbb{R}$ is σ-strongly convex w.r.t. a norm $\|\cdot\|$ if and only if the Fenchel conjugate $\mathcal{R}^\star$ is $\frac{1}{\sigma}$-smooth w.r.t. the dual norm $\|\cdot\|_*$.*

Using the OFTRL update rule and first-order optimality conditions[3], we can write:

$$x_i^{t+1} = \nabla\mathcal{R}^\star\left(\eta\left(\sum_{\tau=1}^{t} u_i^\tau + m_i^{t+1}\right)\right), \quad x_i^t = \nabla\mathcal{R}^\star\left(\eta\left(\sum_{\tau=1}^{t-1} u_i^\tau + m_i^t\right)\right). \tag{6}$$

Using Lemma 9, we have:

$$\|x_i^t - x_i^{t-1}\| \leq \eta\|u_i^t + m_i^{t+1} - m_i^t\|_* = \mathcal{O}(\eta). \tag{7}$$

□

2.4 Near-optimal regret in games with nonnegative regrets

The following result implies that the regret of OFTRL in 2-player zero-sum games is constant (recall for this from the lecture on zero-sum games that the sum of regrets is proportional to the duality gap evaluated at the average iterates which is always nonnegative).

Proposition 10. *If $\sum_{i=1}^N R_i^T \geq 0$, then $R_i^T = \mathcal{O}(1)$.*

Proof. First, we have the RVU bound from Theorem 5:

$$R_i^T \leq \frac{\alpha}{\eta} + \beta\eta\sum_{t=1}^{T}\|u_i^t - u_i^{t-1}\|_*^2 - \frac{\gamma}{\eta}\sum_{t=1}^{T}\|x_i^t - x_i^{t-1}\|^2 \tag{8}$$

$$\leq \frac{\alpha}{\eta} + \beta\eta(N-1)\sum_{j\neq i}\sum_{t=1}^{T}\|x_j^t - x_j^{t-1}\|^2 - \frac{\gamma}{\eta}\sum_{t=1}^{T}\|x_i^t - x_i^{t-1}\|^2. \tag{9}$$

[2] We provide an alternative proof (using the dual of the regularizer) to the proof of Lemma 20 in Syrgkanis et al. (2015).

[3] Obtaining these closed forms also typically requires conditions on the regularizer that we do not expand on here for now.

Summing up the above inequality and using the nonnegativity of the sum if regrets of all players, we have:

$$0 \leq \sum_{i=1}^{N} R_i^T \leq \frac{\alpha N}{\eta} + \beta\eta N(N-1) \sum_{i=1}^{N}\sum_{t=1}^{T} \|x_j^t - x_j^{t-1}\|^2 - \frac{\gamma}{\eta}\sum_{t=1}^{T}\|x_i^t - x_i^{t-1}\|^2$$
$$\leq \frac{\alpha N}{\eta} + \left(\beta N(N-1)\eta - \frac{\gamma}{\eta}\right) \sum_{i=1}^{N}\sum_{t=1}^{T} \|x_j^t - x_j^{t-1}\|^2 . \tag{10}$$

Choosing the step size η as a constant which is sufficiently small, we obtain that the sum of regrets is constant, i.e. $\sum_{i=1}^{N} R_i^T = \mathcal{O}(1)$ (note that this improves the $1/\sqrt{T}$ rate we established for zero-sum games to $1/T$), and that $\sum_{i=1}^{N}\sum_{t=1}^{T}\|x_j^t - x_j^{t-1}\|^2 = \mathcal{O}(1)$ by rearranging the above inequality. Then plugging back the latter result in (8), we obtain that $R_i^T = \mathcal{O}(1)$, which concludes the proof. □

Remark 11. The sum of regrets is also nonnegative for polymatrix zero-sum games.

2.5 Optimal regret in potential games

Proposition 12 (Optimal regret for potential games, Theorem 4.6, Anagnostides et al. (2022b))**.** *Suppose that each player uses OMWU with a sufficiently small learning rate $\eta > 0$. Then, the regret of each player $i \in \mathcal{I}$ is such that $R_i^T = \mathcal{O}(1)$.*

2.6 Regret in general games

Proposition 13. *If all players use OFTRL[4] in a multi-player setting with learning rate $\eta = \mathcal{O}(T^{-\frac{1}{4}})$, the regret of each player is bounded as $\mathcal{O}(T^{\frac{1}{4}})$.*

Proof. Using the RVU bound (Theorem 5), together with Lemma 7 and Lemma 8, we obtain

$$R_i^T \leq \frac{\alpha}{\eta} + \beta\eta \sum_{t=1}^{T}\|u_i^t - u_i^{t-1}\|_*^2 - \frac{\gamma}{\eta}\sum_{t=1}^{T}\|x_i^t - x_i^{t-1}\|^2 \leq \mathcal{O}\left(\frac{1}{\eta}\right) + \mathcal{O}(\eta^3 T) . \tag{11}$$

Selecting the step size to minimize the right-hand side concludes the proof. □

Remark 14. Note that the classical bound in the adversarial setting stemming from (2) is of the form $\mathcal{O}\left(\frac{1}{\eta}\right) + \mathcal{O}(\eta T)$ which results in $\mathcal{O}\left(\sqrt{T}\right)$ bound.

The above bound has been significantly improved in the literature by Daskalakis et al. (2021) to obtain a near-optimal regret bound of $\frac{\text{polylog}(T)}{T}$. Swap regret bounds (which imply external regret bound) have also been improved and actively studied recently (see e.g. Anagnostides et al. (2022a)). In particular, logarithmic regret (which is near-optimal) can be achieved. Showing constant regret is an open question.

[4] or Optimistic Mirror Descent

3 Social Welfare of No-Regret Dynamics

Equilibria are not unique in general. Equilibrium selection is a central question in game theory (Harsanyi and Selten, 1988). This question relates to *efficiency* of equilibria compared to the optimum of a performance metric on the global behavior of the system of N players.

3.1 Social welfare

In the following we mainly focus on the social welfare $W : \mathcal{X} \to \mathbb{R}$ (where $\mathcal{X} := \prod_{j=1}^{N} \mathcal{X}_j$) which we define for every $a \in \mathcal{A}, x \in \mathcal{X}$ as follows:

$$W(a) := \sum_{j=1}^{N} u_j(a)\,, \quad W(x) = \mathbb{E}_{a\sim x}[W(a)]\,. \tag{12}$$

The optimal social welfare is then defined as follows:

$$\mathrm{OPT} := \max_{a=(a_1,\ldots,a_N)\in\mathcal{A}} W(a_1,\ldots,a_N)\,. \tag{13}$$

This quantity corresponds to the optimal welfare achievable in the absence of player incentives and if a central coordinator could dictate each player's strategy.

3.2 Smooth games

The following class of games introduced by Roughgarden (2015) guarantees that if each player unilaterally selects a welfare-optimal strategy, then players collectively guarantee a fraction of the optimal social welfare, irrespective of other players' strategies.

Definition 15 (Smooth games, Roughgarden (2015))**.** *For any $\lambda, \mu \geq 0$, a game is said to be (λ, μ)-smooth with respect to a welfare-optimal strategy profile $x^\star = (x_i^\star, x_{-i}^\star)$ if:*

$$\forall x \in \mathcal{X}, \quad \sum_{i=1}^{N} u_i(x_i^\star, x_{-i}) \geq \lambda OPT - \mu W(x)\,. \tag{14}$$

Roughgarden (2015) shows that some games such as the atomic selfish routing games and locations games are indeed smooth.

For recent progress on social welfare in games using generalized smooth games, see e.g. Chandan et al. (2019). Some recent works also discuss the fragility of the PoA concept (Seaton and Brown, 2023).

3.3 Price of Anarchy bounds for smooth games

Denote by $\mathrm{NE}(\Gamma)$ the set of Nash equilibria of the game Γ.

Proposition 16. *For any (λ, μ)-smooth game Γ, we have*

$$\forall x \in NE(\Gamma), \quad W(x) \geq \frac{\lambda}{1+\mu} OPT\,. \tag{15}$$

The factor $\rho := (1+\mu)/\lambda$ is called the price of anarchy (PoA).

Proof. Let $x^\star \in \operatorname{argmax}_{a\in\mathcal{A}} W(a)$ and let x^{NE} be a Nash equilibrium of the game. Then,

$$W(x^{\mathrm{NE}}) = \sum_{i=1}^{N} u_i(x^{\mathrm{NE}}) \geq \sum_{i=1}^{N} u_i(x_i^\star, x_{-i}^{\mathrm{NE}}) \geq \lambda \mathrm{OPT} - \mu W(x^{\mathrm{NE}})\,, \tag{16}$$

where the first inequality uses the definition of Nash equilibrium and the second uses the definition of a smooth game. Rearranging the inequality concludes the proof. □

The above bound of Proposition 16 can be extended to CCEs using a similar proof upon using the definition of CCE instead of Nash equilibrium.

3.4 Social welfare bounds for no-regret dynamics

The following result relates the (regret) performance of no-regret dynamics to their induced cumulative social welfare, establishing a lower bound of the average welfare of the sequence of play with respect to the optimal welfare of the static game.

Proposition 17 (Proposition 2, Syrgkanis et al. (2015)). *In a (λ, μ)-smooth game, if each player i suffers regret at most R_i^T, then:*

$$\frac{1}{T}\sum_{t=1}^{T} W(x^t) \;\geq\; \frac{\lambda}{1+\mu}\,\mathrm{OPT} \;-\; \frac{1}{1+\mu}\,\frac{1}{T}\sum_{i=1}^{N} R_i^T\,. \tag{17}$$

Proof. By definition of regret, we have for all $i \in \mathcal{I}$ and for all $x_i^\star \in \mathcal{X}_i$,

$$\sum_{t=1}^{T} u_i(x^t) \geq \sum_{t=1}^{T} u_i(x_i^\star, x_{-i}^t) - R_i^T. \tag{18}$$

Summing over all players and using the smoothness property:

$$\begin{aligned}
\sum_{t=1}^{T} W(x^t) &= \sum_{t=1}^{T}\sum_{i=1}^{N} u_i(x^t) \\
&= \sum_{i=1}^{N}\sum_{t=1}^{T} u_i(x^t) \\
&\geq \sum_{i=1}^{N}\sum_{t=1}^{T} u_i(x_i^\star, x_{-i}^t) - R_i^T \\
&= \sum_{t=1}^{T}\sum_{i=1}^{N} u_i(x_i^\star, x_{-i}^t) - \sum_{i=1}^{N} R_i^T \\
&\geq \sum_{t=1}^{T}(\lambda\mathrm{OPT} - \mu W(x^t)) - \sum_{i=1}^{N} R_i^T \\
&= \lambda T\cdot\mathrm{OPT} - \mu\sum_{t=1}^{T} W(x^t) - \sum_{i=1}^{N} R_i^T\,.
\end{aligned}$$

Rearranging the above inequality and dividing by T yields the desired inequality. □

The above proposition shows that we can approximate the optimal welfare using decoupled no-regret dynamics in smooth games. The approximation is driven by the quantity $\frac{1}{1+\mu}\,\frac{1}{T}\sum_{i=1}^{N} R_i^T$. Depending on the no-regret algorithm used, we obtain a rate on the average social welfare. If individual regret is $R_i^T = O\Big(\sqrt{\log(d)\,T}\Big)$, then the average welfare converges to PoA at a rate of $O\Big(n\sqrt{\log(d)/T}\Big)$. Using optimism, this rate can be improved to $O\big(n^2\log(d)/T\big)$.

References

Ioannis Anagnostides. Computational game solving cmu course, guest lecture 9. https://www.cs.cmu.edu/~sandholm/cs15-888F24/Lecture_9_Ioannis_Anagnostides_guest.pdf, 2024.

Ioannis Anagnostides, Gabriele Farina, Christian Kroer, Chung-Wei Lee, Haipeng Luo, and Tuomas Sandholm. Uncoupled learning dynamics with $o(\log t)$ swap regret in multiplayer games. *Advances in Neural Information Processing Systems*, 35:3292–3304, 2022a.

Ioannis Anagnostides, Ioannis Panageas, Gabriele Farina, and Tuomas Sandholm. On last-iterate convergence beyond zero-sum games. In *Proceedings of the 39th International Conference on Machine Learning*, pages 536–581, 2022b. URL https://proceedings.mlr.press/v162/anagnostides22a.html.

Rahul Chandan, Dario Paccagnan, and Jason R Marden. When smoothness is not enough: Toward exact quantification and optimization of the price-of-anarchy. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 4041–4046. IEEE, 2019.

Xi Chen and Binghui Peng. Hedging in games: Faster convergence of external and swap regrets. *Advances in Neural Information Processing Systems*, 33:18990–18999, 2020.

Chao-Kai Chiang, Tianbao Yang, Chia-Jung Lee, Mehrdad Mahdavi, Chi-Jen Lu, Rong Jin, and Shenghuo Zhu. Online optimization with gradual variations. In *Conference on Learning Theory*, pages 6–1. JMLR Workshop and Conference Proceedings, 2012.

Constantinos Daskalakis, Maxwell Fishelson, and Noah Golowich. Near-optimal no-regret learning in general games. *Advances in Neural Information Processing Systems*, 34:27604–27616, 2021.

John C Harsanyi and Reinhard Selten. A general theory of equilibrium selection in games. *MIT Press Books*, 1, 1988.

Alexander Rakhlin and Karthik Sridharan. Online learning with predictable sequences. In *Conference on Learning Theory*, pages 993–1019. PMLR, 2013.

Tim Roughgarden. Intrinsic robustness of the price of anarchy. *Journal of the ACM (JACM)*, 62 (5):1–42, 2015.

Joshua H. Seaton and Philip N. Brown. On the intrinsic fragility of the price of anarchy. *IEEE Control Systems Letters*, 7:3573–3578, 2023. doi: 10.1109/LCSYS.2023.3335315.

Vasilis Syrgkanis, Alekh Agarwal, Haipeng Luo, and Robert E Schapire. Fast convergence of regularized learning in games. *Advances in Neural Information Processing Systems*, 28, 2015.