Economics, AI, and Optimization
Lecture Note 5 Learning Equilibrium via Regret Minimization

Christian Kroer*

February 1, 2022

1 Recap

We have covered a slew of no-regret algorithms: hedge, online mirror descent (OMD), regret matching (RM), and RM$^+$. All of these algorithms can be used for the case of solving two-player zero-sum matrix games of the form $\min_{x\in\Delta^n}\max_{y\in\Delta^m}\langle x, Ay\rangle$. In this lecture note we will cover how to compute a saddle point of the more general case of

$$\min_{x\in X}\max_{y\in Y} f(x,y)$$

where f is convex-concave, meaning that $f(\cdot, y)$ is convex for all fixed y, and $f(x,\cdot)$ is concave for all fixed x, and lower/upper semicontinuous. We will then look at some experiments on practical performance for the matrix-game case. We will also compare to an algorithm that have stronger theoretical guarantees.

2 From Regret to Nash Equilibrium

In order to use these algorithms for computing Nash equilibrium, we will run a repeated game between the x and y players. We will assume that each player has access to some regret-minimizing algorithm R_x and R_y (we will be a bit loose with notation here and implicitly assume that R_x and R_y keep a state that may depend on the sequence of losses and decisions) The game is as follows:

- Initialize x_1, y_1 to be uniform distributions over actions
- At time t, let x_t be the recommendation from R_x and y_t be the recommendation from R_y
- Let R_x and R_y observe losses $g_t = f(\cdot, y_t), \ell_t = f(x_t, \cdot)$ respectively

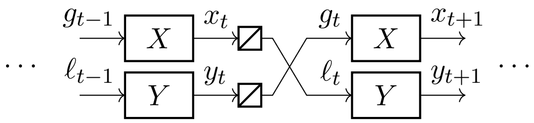

Figure 1: The flow of strategies and losses in regret minimization for games.

[*]Department of Industrial Engineering and Operations Research, Columbia University. Email: christian.kroer@columbia.edu.

For a strategy pair $\bar{x}, \bar{y}$, we will measure proximity to Nash equilibrium via the *saddle-point residual* (SPR):

$$\xi(\bar{x}, \bar{y}) := \left[\max_{y \in Y} f(\bar{x}, y) - f(\bar{x}, \bar{y})\right] + \left[f(\bar{x}, \bar{y}) - \min_{x \in X} f(x, \bar{y})\right] = \max_{y \in Y} f(\bar{x}, y) - \min_{x \in X} f(x, \bar{y}).$$

Each bracketed term represents how much each player can improve by deviating from $\bar{y}$ or $\bar{x}$ respectively, given the strategy profile $(\bar{x}, \bar{y})$. In game-theoretic terms the brackets are how much each player improves by best responding.

Now, suppose that the regret-minimizing algorithms guarantee regret bounds of the form

$$\begin{aligned} \max_{y \in Y} \sum_{t=1}^{T} f(x_t, y) - \sum_{t=1}^{T} f(x_t, y_t) \leq \epsilon_y \\ \sum_{t=1}^{T} f(x_t, y_t) - \min_{x \in X} \sum_{t=1}^{T} f(x, y_t) \leq \epsilon_x, \end{aligned} \tag{1}$$

then the following folk theorem holds

Theorem 1. *Suppose* (1) *holds, then for the average strategies* $\bar{x} = \frac{1}{T}\sum_{t=1}^{T} x_t, \bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$ *the SPR is bounded by*

$$\xi(\bar{x}, \bar{y}) \leq \frac{(\epsilon_x + \epsilon_y)}{T}.$$

Proof. Summing the two inequalities in (1) we get

$$\begin{aligned} \epsilon_x + \epsilon_y \geq & \max_{y \in Y} \sum_{t=1}^{T} f(x_t, y) - \sum_{t=1}^{T} f(x_t, y_t) + \sum_{t=1}^{T} f(x_t, y_t) - \min_{x \in X} \sum_{t=1}^{T} f(x, y_t) \\ = & \max_{y \in Y} \sum_{t=1}^{T} f(x_t, y) - \min_{x \in X} \sum_{t=1}^{T} f(x, y_t) \\ \geq & T \left[\max_{y \in Y} f(\bar{x}, y) - \min_{x \in X} f(x, \bar{y})\right], \end{aligned}$$

where the inequality is by f being convex-concave.

□

So now we know how to compute a Nash equilibrium: simply run the above repeated game with each player using a regret-minimizing algorithm, and the uniform average of the strategies will converge to a Nash equilibrium.

Figure 2 shows the performance of the regret-minimization algorithms taught so far in the course, when used to compute a Nash equilibrium of a zero-sum matrix game via Theorem 1. Performance is shown on 3 randomized matrix game classes where entries in A are sampled according to: 100-by-100 uniform $[0, 1]$, 500-by-100 standard Gaussian, and 100-by-100 standard Gaussian. All plots are averaged across 50 game samples per setup. We show one addition algorithm for reference: the *mirror prox* algorithm, which is an offline optimization algorithm that converges to a Nash equilibrium at a rate of $O\left(\frac{1}{T}\right)$. It's an accelerated variant of mirror descent, and it similarly relies on a distance-generating function d. The plot shows mirror prox with the Euclidean distance.

As we see in Figure 2, mirror prox indeed performs better than all the $O\left(\frac{1}{\sqrt{T}}\right)$ regret minimizers using the setup for Theorem 1. On the other hand, the entropy-based variant of OMD, which has a $\log n$ dependence on the dimension n, performs much worse than the algorithms with $\sqrt{n}$ dependence.

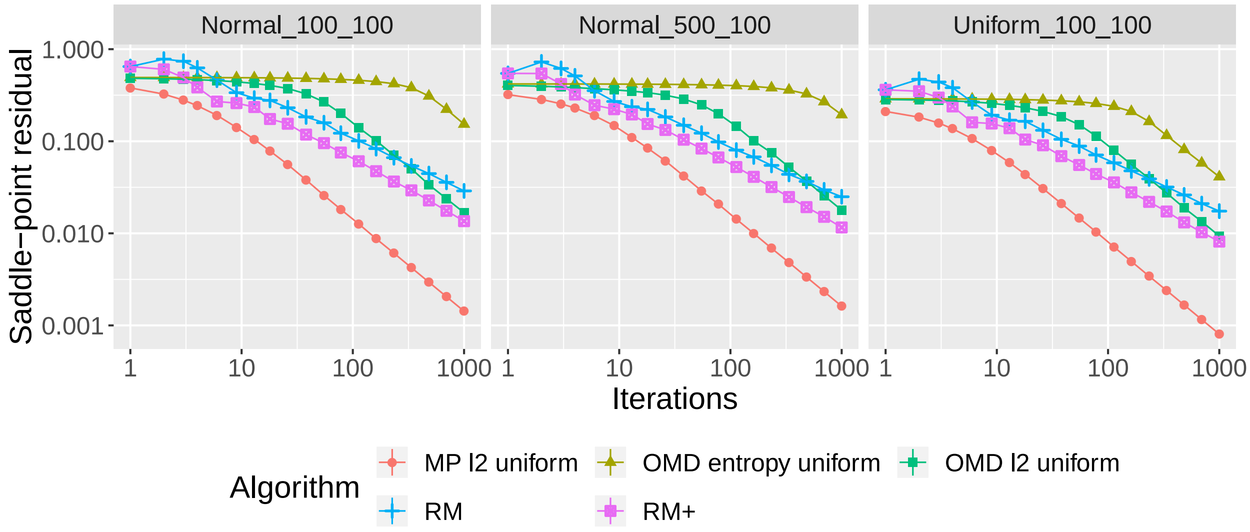

Figure 2: Plots showing the performance of four different regret-minimization algorithms for computing Nash equilibrium, all using Theorem 1. Mirror prox with uniform averaging is also shown as a reference point.

3 Alternation

Let's try making a small tweak now. We will consider what is usually called *alternation*. In alternation, the players are no longer symmetric: one player sees the loss based on the previous strategy of the other player as before, but the second player sees the loss associated to the current strategy.

- Initialize x_1, y_1 to be uniform distributions over actions
- At time t, let x_t be the recommendation from R_x
- The y player observes loss $f(x_t, \cdot)$
- y_t is the recommendation from R_y after observing $f(x_t, \cdot)$
- The x player observes loss $f(\cdot, y_t)$

Suppose that the regret-minimizing algorithms guarantee regret bounds of the form

$$\begin{aligned} \max_{y \in Y} \sum_{t=1}^T f(x_{t+1}, y) - \sum_{t=1}^T f(x_{t+1}, y_t) \leq \epsilon_y \\ \sum_{t=1}^T f(x_t, y_t) - \min_{x \in X} \sum_{t=1}^T f(x, y_t) \leq \epsilon_x. \end{aligned} \tag{2}$$

Theorem 2. *Suppose we run two regret minimizer with alternation and they give the guarantees in (2). Then the average strategies* $\bar{x} = \frac{1}{T}\sum_{t=1}^T x_{t+1}, \bar{y} = \frac{1}{T}\sum_{t=1}^T y_t$.

$$\xi(\bar{x}, \bar{y}) \leq \frac{\epsilon_x + \epsilon_y + \sum_{t=1}^T \left(f(x_{t+1}, y_t) - f(x_t, y_t)\right)}{T}$$

Proof. As before we sum the regret bounds to get

$$\begin{aligned}\epsilon_x + \epsilon_y \geq& \max_{y\in Y} \sum_{t=1}^{T} f(x_{t+1}, y) - \sum_{t=1}^{T} f(x_{t+1}, y_t) + \sum_{t=1}^{T} f(x_t, y_t) - \min_{x\in X} \sum_{t=1}^{T} f(x, y_t)\\ =& \max_{y\in Y} \sum_{t=1}^{T} f(x_{t+1}, y) - \min_{x\in X} \sum_{t=1}^{T} f(x, y_t) - \sum_{t=1}^{T} [f(x_{t+1}, y_t) - f(x_t, y_t)]\\ \geq& T\left[\max_{y\in Y} f(\bar{x}, y) - \min_{x\in X} f(x, \bar{y})\right] - \sum_{t=1}^{T} [f(x_{t+1}, y_t) - f(x_t, y_t)]\end{aligned}$$

□

Figure 3 shows the performance of the same set of regret-minimization algorithms but now using the setup from Theorem 2. Mirror prox is shown exactly as before.

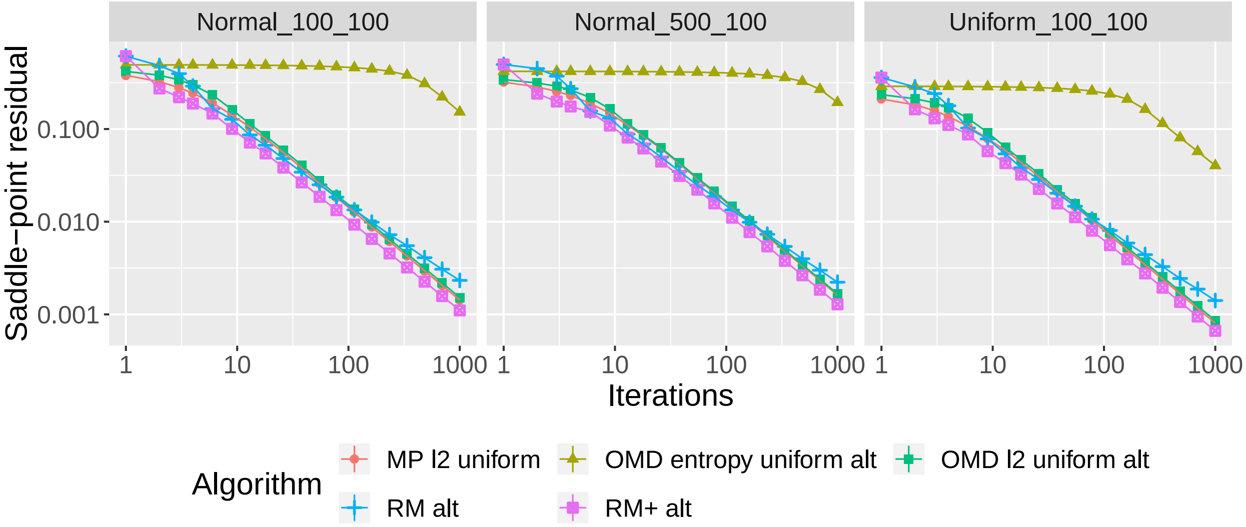

Figure 3: Plots showing the performance of four different regret-minimization algorithms for computing Nash equilibrium, all using Theorem 2. Mirror prox with uniform averaging is also shown as a reference point.

Amazingly, Figure 3 shows that with alternation, OMD with Euclidean DGF, regret matching, and RM$^+$ all performs about on par with mirror prox.

4 Increasing Iterate Averaging

Now we will look at one final tweak. In Theorems 1 and 2 we generated a solution by uniformly averaging iterates. We will now consider polynomial averaging schemes of the form

$$\bar{x} = \frac{1}{\sum_{t=1}^{T} t^q} \sum_{t=1}^{T} t^q x_t, \quad \bar{y} = \frac{1}{\sum_{t=1}^{T} t^q} \sum_{t=1}^{T} t^q y_t.$$

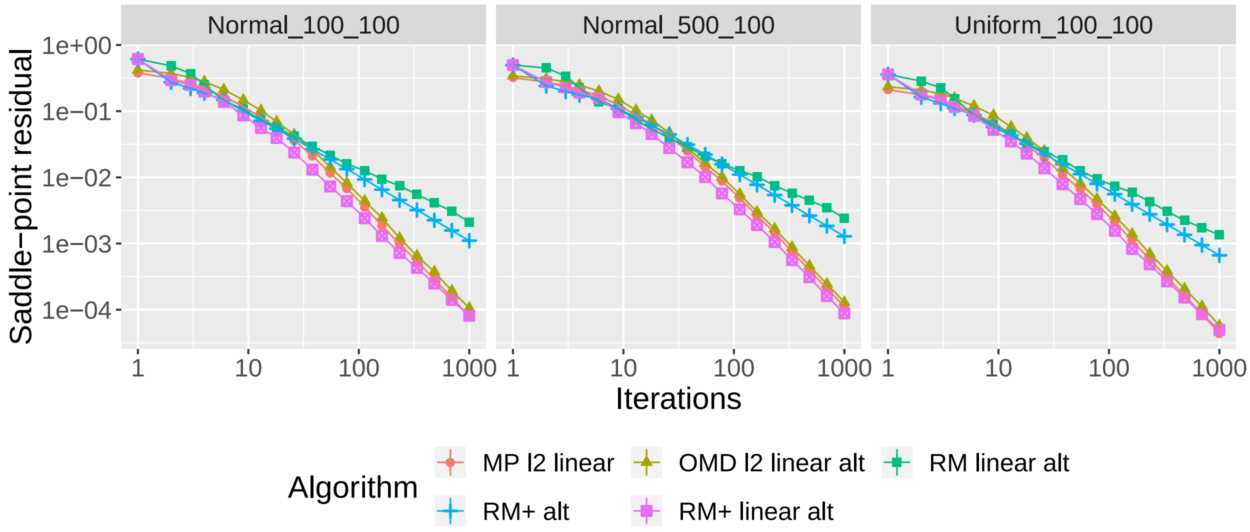

Figure 4: Plots showing the performance of four different regret-minimization algorithms for computing Nash equilibrium, all using Theorem 2. All algorithms use linear averaging. RM$^+$ with uniform averaging is shown as a reference point.

Figure 4 shows the performance of the same set of regret-minimization algorithms but now using the setup from Theorem 2 and linear averaging in all algorithms, including mirror prox. The fastest algorithm with uniform averaging, RM$^+$ with alternation, is shown for reference. OMD with Euclidean DGF and RM$^+$ with alternation both gain another order of magnitude in performance by introducing linear averaging.

It can be shown that RM$^+$, online mirror descent, and mirror prox, all work with polynomial averaging schemes.

5 Historical Notes and Further Reading

The derivation of a folk theorem for alternation in matrix games was by Burch et al. [2], after Farina et al. [4] pointed out that the original folk theorem does not apply when using alternation. The general convex-concave case is new, although easily derived from the existing results.

The fact that Euclidean distance seems to perform better than entropy when solving matrix games in practice has been observed in a few different algorithms both first-order methods [3, 6] and regret-minimization algorithms [5]. The fact that OMD with Euclidean distance performs much better after adding alternation has not been observed before.

Results for polynomial averaging schemes were shown by Tammelin et al. [7], Brown and Sandholm [1] for RM$^+$, in Nemirovski's lecture notes at `https://www2.isye.gatech.edu/~nemirovs/LMCO_LN2019NoSolutions.pdf` for mirror descent and mirror prox, and for several other primal-dual first-order methods by Gao et al. [6].

References

[1] 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.

[2] Neil Burch, Matej Moravcik, and Martin Schmid. Revisiting cfr+ and alternating updates. *Journal of Artificial Intelligence Research*, 64:429–443, 2019.

[3] Antonin Chambolle and Thomas Pock. On the ergodic convergence rates of a first-order primal–dual algorithm. *Mathematical Programming*, 159(1-2):253–287, 2016.

[4] Gabriele Farina, Christian Kroer, and Tuomas Sandholm. Online convex optimization for sequential decision processes and extensive-form games. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 1917–1925, 2019.

[5] Gabriele Farina, Christian Kroer, and Tuomas Sandholm. Optimistic regret minimization for extensive-form games via dilated distance-generating functions. In *Advances in Neural Information Processing Systems*, pages 5222–5232, 2019.

[6] Yuan Gao, Christian Kroer, and Donald Goldfarb. Increasing iterate averaging for solving saddle-point problems. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021.

[7] Oskari Tammelin, Neil Burch, Michael Johanson, and Michael Bowling. Solving heads-up limit Texas hold'em. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*, 2015.