MS&E 233
Game Theory, Data Science and AI
Lecture 3

Vasilis Syrgkanis

Assistant Professor

Management Science and Engineering

(by courtesy) Computer Science and Electrical Engineering

Institute for Computational and Mathematical Engineering

Class Music Auction!

We will be experimenting with putting music for the first three minutes of the class as people arrive!

You have the chance to choose the song of the day!

Each of you has a total budget of 100 fake dollars for the whole class! You can choose to spend them however you want on each lecture.

For each lecture you can choose to bid anywhere from 0 to 20 dollars.

We will then choose uniformly at random among the highest bidders. The winner of the auction will get to choose the song of the day and they have to pay their bid, i.e. the amount they bid will be subtracted from their 100$ budget.

If you submit an illegal bid (i.e. a bid that goes beyond your total budget, your bid will be disqualified and ignored).

Please be appropriate in your choice of songs; I might need to censor and ask you to choose something else. I'll be emailing the winner on the morning of the lecture to email me the spotify link for the song.

Submit your bid by 11:59pm the day before the lecture. You should submit your bid using the corresponding canvas quiz that will be setup for each lecture

Class Music Auction: Game Theory, Data Science and AI (stanford.edu)

Go to canvas and check the quizzes section.

If there is no participation in the auction, I'll just choose the music myself. But that's not much fun...

Spotify playlist that will be populated with the songs we play each day: https://open.spotify.com/playlist/03yGb6URnCzG4pVV6RhK4C?si=wpINDMSGRJOho_6daaSLsA&pt=ff706933952e0f64d8f8b797368a83ed

Computational Game Theory for Complex Games

1
- Basics of game theory and zero-sum games (T)
- Basics of online learning theory (T)
- **Solving zero-sum games via online learning (T)**
- *HW1: implement simple algorithms to solve zero-sum games*
- Applications to ML and AI (T+A)
- *HW2: implement boosting as solving a zero-sum game*

2
- Basics and applications of extensive-form games (T+A)
- Solving extensive-form games via online learning (T)
- *HW3: implement agents to solve very simple variants of poker*

3
- General games and equilibria (T)
- Online learning in general games, multi-agent RL (T+A)
- *HW4: implement no-regret algorithms that converge to correlated equilibria in general games*

Data Science for Auctions and Mechanisms

4
- Basics and applications of auction theory (T+A)
- Learning to bid in auctions via online learning (T)
- *HW5: implement bandit algorithms to bid in ad auctions*

5
- Optimal auctions and mechanisms (T)
- Simple vs optimal mechanisms (T)
- *HW6: calculate equilibria in simple auctions, implement simple and optimal auctions, analyze revenue empirically*

6
- Optimizing mechanisms from samples (T)
- Online optimization of auctions and mechanisms (T)
- *HW7: implement procedures to learn approximately optimal auctions from historical samples and in an online manner*

Further Topics

7
- Econometrics in games and auctions (T+A)
- A/B testing in markets (T+A)
- *HW8: implement procedure to estimate values from bids in an auction, empirically analyze inaccuracy of A/B tests in markets*

Guest Lectures

- Mechanism Design for LLMs, Renato Paes Leme, Google Research
- Auto-bidding in Sponsored Search Auctions, Kshipra Bhawalkar, Google Research

Learning and Zero-Sum Games

Reminder: Two Player Zero-Sum Games

- Player one (“min” player or “row” player)
- Player two (“max” player or “column” player)
- Player one has n possible actions
- Player two has m possible actions

- If player one chooses action i and player two chooses action j then player one incurs loss $A[i, j]$ and player two gains utility $A[i, j]$

Reminder: Equilibrium via Min-Max Theorem

- Suppose that both players behave *pessimistically*
- Row (min) player thinks: "I'll choose a strategy x such that I'll try to minimize the worst-case loss that the other player can cause me"

$$\bar{x} = \underset{x}{\operatorname{argmin}} \left(\max_{y} x' A y \right)$$

- Column (max) player thinks: "I'll choose a strategy y such that I'll try to maximize the worst-case utility that the other player will allow me to get"

$$\bar{y} = \underset{y}{\operatorname{argmax}} \left(\min_{x} x' A y \right)$$

Reminder: Equilibrium via Min-Max Theorem

- Suppose both players behave pessimistically

$$\bar{x} = \operatorname{argmin}_x \left(\max_y x' A y \right), \qquad \bar{y} = \operatorname{argmax}_y \left(\min_x x' A y \right)$$

- **Von Neuman's Min-Max Theorem [1928]:** Pessimistic value that each player achieves is the same

$$\min_x \max_y x' A y = \max_y \min_x x' A y$$

Smallest loss that min player can achieve if max chooses $\bar{y}$

$$\boxed{\bar{x}' A \bar{y}} \le \boxed{\max_y \bar{x}' A y} = \boxed{\min_x \max_y x' A y} = \boxed{\max_y \min_x x' A y} = \boxed{\min_x x' A \bar{y}}$$

Loss of min player at $(\bar{x}, \bar{y})$

Pessimistic loss if I choose $\bar{x}$

Best pessimistic loss by definition of $\bar{x}$

Best pessimistic utility that max player can achieve

Pessimistic utility that max player achieves by using $\bar{y}$

Equilibrium via Learning

- What if we have the players play the game repeatedly?
- At each period t each player picks a choice distribution, (x_t, y_t)

Are there dynamics that will lead to a mixed Nash equilibrium?

What if each player uses a no-regret algorithm!

Equilibrium via No-Regret Learning

- Think of the problem that the x-player faces:
  - At each period t, pick a choice distribution x_t
  - Incur loss $x_t^\top A y_t$ and observe loss each action would incur: $A y_t$
  - *Incur loss* $x_t^\top \ell_t$ *and observe loss each action would incur:* $\ell_t \coloneqq A y_t$

Equilibrium via No-Regret Learning

- Think of the problem that the x-player faces:
  - At each period t, pick a choice distribution x_t
  - Incur loss $x_t^\top A y_t$ and observe loss each action would incur: $A y_t$
  - *Incur loss $x_t^\top \ell_t$ and observe loss each action would incur: $\ell_t \coloneqq A y_t$*
- Think of the problem the y-player faces
  - At each period t, pick a choice distribution y_t
  - Incur loss $-x_t^\top A y_t$ and observe loss each action would incur: $-A^\top x_t$
  - *Incur loss $\tilde{\ell}_t^\top y_t$ and observe loss each action would incur: $\tilde{\ell}_t \coloneqq -A^\top x_t$*
- Both players face a no-regret learning problem!

No-Regret Implications

- We now know how to construct no-regret algorithms! (e.g. EXP)

$$x_t \propto x_{t-1} \exp(-\eta\, \ell_{t-1}), \qquad y_t \propto y_{t-1} \exp\left(-\eta \tilde{\ell}_{t-1}\right)$$

- What this implies is that in the limit as $T \to \infty$ for some $\epsilon \to 0$

$$\boxed{\frac{1}{T}\sum_{t=1}^{T} x_t^\top A y_t} \le \boxed{\min_x \frac{1}{T}\sum_{t=1}^{T} x^\top A y_t} + \boxed{\epsilon} = \min_x x^\top A \boxed{\left(\frac{1}{T}\sum_{t=1}^{T} y_t\right)} + \epsilon$$

Expected average loss of x-player

Average loss of x-player's best fixed choice distribution in hindsight

Regret

Average choice distribution of y-player

No-Regret Implications

- We now know how to construct no-regret algorithms! (e.g. EXP)

$$x_t \propto x_{t-1} \exp(-\eta\, \ell_{t-1}), \qquad y_t \propto y_{t-1} \exp\left(-\eta \tilde{\ell}_{t-1}\right)$$

- What this implies is that in the limit as $T \to \infty$ for some $\epsilon \to 0$

$$\frac{1}{T}\sum_{t=1}^{T} x_t^\top A y_t \le \min_x \frac{1}{T}\sum_{t=1}^{T} x^\top A y_t + \epsilon = \min_x x^\top A \left(\frac{1}{T}\sum_{t=1}^{T} y_t\right) + \epsilon$$

$$\underbrace{\frac{1}{T}\sum_{t=1}^{T} x_t^\top A y_t}_{\text{Expected average utility of } y\text{-player}} \ge \underbrace{\max_y \frac{1}{T}\sum_{t=1}^{T} x_t^\top A y}_{\text{Average utility of } y\text{-player's best fixed choice distribution in hindsight}} - \underbrace{\epsilon}_{\text{Regret}} = \max_y \underbrace{\left(\frac{1}{T}\sum_{t=1}^{T} x_t^\top\right)}_{\text{Average choice distribution of } x\text{-player}} A y - \epsilon$$

No-Regret Implications

- We now know how to construct no-regret algorithms! (e.g. EXP)

$$x_t \propto x_{t-1} \exp(-\eta\, \ell_{t-1}), \qquad y_t \propto y_{t-1} \exp\left(-\eta \tilde{\ell}_{t-1}\right)$$

- What this implies is that in the limit as $T \to \infty$ for some $\epsilon \to 0$
- Define the *average choice distributions* as $\bar{x}, \bar{y}$

$$\bar{x} \coloneqq \frac{1}{T}\sum_{t=1}^{T} x_t, \quad \text{then} \quad \frac{1}{T}\sum_{t=1}^{T} x_t^\top A y_t \le \min_x x^\top A \bar{y} + \epsilon$$

Expected average loss of x-player is a “best-response” to average strategy $\bar{y}$ of y-player

$$\bar{y} \coloneqq \frac{1}{T}\sum_{t=1}^{T} y_t, \quad \text{then} \quad \frac{1}{T}\sum_{t=1}^{T} x_t^\top A y_t \ge \max_y \bar{x} A y - \epsilon$$

Expected average utility of y-player is a “best-response” to average strategy $\bar{x}$ of x-player

Candidate Equilibrium

- x-player's average loss is a best-response to $\bar{y}$
- y-player's average utility is a best-response to $\bar{x}$
- Could it be that maybe $(\bar{x}, \bar{y})$ is an equilibrium?

Average loss (utility) $\boxed{\frac{1}{T}\sum_t x_t^\top A y_t} \neq \boxed{\bar{x}^\top A \bar{y}}$ Loss (utility) under average strategies

- We need to see if loss (utility) under average strategies also satisfies the same best-response property
- **Crucial:** Average loss of x-player $=$ Average utility of y-player

No-Regret Implications

- Define the *average choice distributions* as $\bar{x}, \bar{y}$

$$\frac{1}{T}\sum_{t=1}^{T} x_t^\top A y_t \leq \min_x x^\top A \bar{y} + \epsilon$$

$$\|$$

$$\frac{1}{T}\sum_{t=1}^{T} x_t^\top A y_t \geq \max_y \bar{x}^\top A y - \epsilon$$

Average loss of x-player =
Average utility of y-player

No-Regret Implications

- Define the *average choice distributions* as $\bar{x}, \bar{y}$

$$\frac{1}{T}\sum_{t=1}^{T} x_t^\top A y_t \le \min_x x^\top A \bar{y} + \epsilon \le \bar{x}^\top A \bar{y} + \epsilon$$

$$\|$$

$$\frac{1}{T}\sum_{t=1}^{T} x_t^\top A y_t \ge \max_y \bar{x}^\top A y - \epsilon \ge \bar{x}^\top A \bar{y} - \epsilon$$

Average loss of x-player =
Average utility of y-player

Loss (utility) of
average strategies

No-Regret Implications

- Define the *average choice distributions* as $\bar{x}, \bar{y}$

$$\bar{x}^\top A \bar{y} \geq \max_y \bar{x}^\top A y - 2\epsilon$$

$$\frac{1}{T}\sum_{t=1}^{T} x_t^\top A y_t \leq \min_x x^\top A \bar{y} + \epsilon \leq \bar{x}^\top A \bar{y} + \epsilon$$

$$\|$$

$$\frac{1}{T}\sum_{t=1}^{T} x_t^\top A y_t \geq \max_y \bar{x}^\top A y - \epsilon \geq \bar{x}^\top A \bar{y} - \epsilon$$

Average loss of x-player = Average utility of y-player

$\leq \geq$

Loss (utility) of average strategies

No-Regret Implications

- Define the *average choice distributions* as $\bar{x}, \bar{y}$

$$\bar{x}^\top A \bar{y} \geq \max_y \bar{x}^\top A y - 2\epsilon$$

$$\bar{x}^\top A \bar{y} \leq \min_x x^\top A \bar{y} + 2\epsilon$$

$$\frac{1}{T}\sum_{t=1}^{T} x_t^\top A y_t \leq \min_x x^\top A \bar{y} + \epsilon \leq \bar{x}^\top A \bar{y} + \epsilon$$

$$\|$$

$$\frac{1}{T}\sum_{t=1}^{T} x_t^\top A y_t \geq \max_y \bar{x}^\top A y - \epsilon \geq \bar{x}^\top A \bar{y} - \epsilon$$

Average loss of x-player = Average utility of y-player

$\leq$ $\geq$

Loss (utility) of average strategies

No-Regret Implications

- Define the *average choice distributions* as $\bar{x}, \bar{y}$

$$\bar{x}^\top A \bar{y} \geq \max_y \bar{x}^\top A y - 2\epsilon$$

$$\bar{x}^\top A \bar{y} \leq \min_x x^\top A \bar{y} + 2\epsilon$$

$(\bar{x}, \bar{y})$ is a 2ϵ-approximate equilibrium

$(\bar{x}, \bar{y}) \to$ equilibrium as $T \to \infty$

$$\frac{1}{T}\sum_{t=1}^{T} x_t^\top A y_t \leq \min_x x^\top A \bar{y} + \epsilon \leq \bar{x}^\top A \bar{y} + \epsilon$$

$$\parallel$$

$$\frac{1}{T}\sum_{t=1}^{T} x_t^\top A y_t \geq \max_y \bar{x}^\top A y - \epsilon \geq \bar{x}^\top A \bar{y} - \epsilon$$

Average loss of x-player = Average utility of y-player

$\leq$ $\geq$

Loss (utility) of average strategies

Main Takeaway: Equilibrium via No-Regret

Theorem. If two players play repeatedly a zero-sum game and each player uses any no-regret algorithm to pick their action distributions (x_t, y_t), then the average action distributions of each player

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

are a 2ϵ-approximate Nash equilibrium (where ϵ is the regret at of each algorithm after T periods). Hence,

$$(\bar{x}, \bar{y}) \to \text{equilibrium as } T \to \infty$$

Main Takeaway: Equilibrium via No-Regret

Corollary. If two players play repeatedly a zero-sum game, with n rows and m columns, and each player uses EXP with step size $\eta = \sqrt{\log\max(n,m)/2T}$, to pick their action distributions (x_t, y_t), then the average action distributions of each player

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

are a 2ϵ- approximate Nash equilibrium, with $\epsilon = \sqrt{\frac{2\log\max(n,m)}{T}}$, i.e.

$$\text{Regret}_x(\bar{x},\bar{y}) := \bar{x}^\top A\bar{y} - \min_x x^\top A\bar{y} \le 2\epsilon$$

$$\text{Regret}_y(\bar{x},\bar{y}) := \max_y \bar{x}^\top Ay - \bar{x}^\top A\bar{y} \le 2\epsilon$$

Minimax Theorem via No-Regret

- Define the *average choice distributions* as $\bar{x}, \bar{y}$

$$\frac{1}{T}\sum_{t=1}^{T} x_t^\top A y_t \le \min_x x^\top A \bar{y} + \epsilon$$

$$\|$$

$$\frac{1}{T}\sum_{t=1}^{T} x_t^\top A y_t \ge \max_y \bar{x}^\top A y - \epsilon$$

Average loss of x-player =
Average utility of y-player

Minimax Theorem via No-Regret

- Define the *average choice distributions* as $\bar{x}, \bar{y}$

$$\frac{1}{T}\sum_{t=1}^{T} x_t^\top A y_t \le \min_x x^\top A \bar{y} + \epsilon \le \max_y \min_x x^\top A y + \epsilon$$

$$\|$$

$$\frac{1}{T}\sum_{t=1}^{T} x_t^\top A y_t \ge \max_y \bar{x}^\top A y - \epsilon \ge \min_x \max_y x^\top A y - \epsilon$$

Average loss of x-player =
Average utility of y-player

Minimax Theorem via No-Regret

- Define the *average choice distributions* as $\bar{x}, \bar{y}$

$$\frac{1}{T}\sum_{t=1}^{T} x_t^\top A y_t \le \min_x x^\top A\bar{y} + \epsilon \le \max_y \min_x x^\top A y + \epsilon$$

$$\|$$

$$\frac{1}{T}\sum_{t=1}^{T} x_t^\top A y_t \ge \max_y \bar{x}^\top A y - \epsilon \ge \min_x \max_y x^\top A y - \epsilon$$

$$\max_y \min_x x^\top A y \ge \min_x \max_y x^\top A y + 2\epsilon$$

Average loss of x-player =
Average utility of y-player

Minimax Theorem via No-Regret

Theorem. Existence of no-regret algorithms implies (as $\epsilon \to 0$) that

$$\max_y \min_x x^\top A y \geq \min_x \max_y x^\top A y$$

The other direction is trivial *(why?)*

$$\max_y \min_x x^\top A y \leq \min_x \max_y x^\top A y$$

Thus

$$\max_y \min_x x^\top A y = \min_x \max_y x^\top A y$$

Wait; we saw no-regret algorithms exist for convex losses too.
What does that imply for games?

Convex-Concave Zero-Sum Games

- Player one (“min” player) chooses a vector x from a convex set $\mathcal{X}$
- Player two (“max” player) chooses a vector y from a convex set $\mathcal{Y}$
- The min player incurs loss $\ell(x, y)$, with $\ell(\cdot, y)$ a convex function
- The max player receives utility $\ell(x, y)$ (*equiv. incurs loss* $-\ell(x, y)$), with $\ell(x,\cdot)$ a concave function (*equiv.* $-\ell(x,\cdot)$ *a convex function*)

- We typically represent this game by its min-max formulation

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

Equilibrium via No-Regret Learning

- Think of the problem that the x-player faces:
  - At each period t, pick a vector x_t from a convex set $\mathcal{X}$
  - Incur loss $\ell(x_t, y_t)$; observe convex loss function: $\ell(\cdot, y_t)$

- Think of the problem the y-player faces
  - At each period t, pick a vector y_t from a convex set $\mathcal{Y}$
  - Incur loss $-\ell(x_t, y_t)$; observe convex loss function: $-\ell(x_t, \cdot)$

- Both players face a convex no-regret learning problem!

Equilibrium via No-Regret Learning

- Think of the problem that the x-player faces:
  - At each period t, pick a vector x_t from a convex set $\mathcal{X}$
  - Incur loss $\ell(x_t, y_t)$; observe convex loss function: $\ell(\cdot, y_t)$

simplex $\Delta(n)$ in the finite action case

$x_t^\top A y_t$ in the finite action case

$A y_t$ in the finite action case

- Think of the problem the y-player faces
  - At each period t, pick a vector y_t from a convex set $\mathcal{Y}$
  - Incur loss $-\ell(x_t, y_t)$; observe convex loss function: $-\ell(x_t, \cdot)$

simplex $\Delta(m)$ in the finite action case

$-x_t^\top A y_t$ in the finite action case

$-A^\top x_t$ in the finite action case

- Both players face a convex no-regret learning problem!

No-Regret Implications

- We know no-regret algorithms exist! (e.g., online gradient descent)

$$x_t = x_{t-1} - \eta \nabla_{\mathrm{x}} \ell(x_{t-1}, y_{t-1}), \qquad y_t = y_{t-1} + \eta \nabla_y \ell(x_{t-1}, y_{t-1})$$

- What this implies is that in the limit as $T \to \infty$ for a regret $\epsilon \to 0$

$$\frac{1}{T}\sum_{t=1}^{T} \ell(x_t, y_t) \le \min_x \frac{1}{T}\sum_{t=1}^{T} \ell(x, y_t) + \epsilon \le \min_x \ell(x, \bar{y}) + \epsilon$$

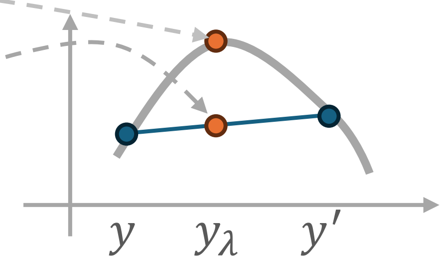
Average loss of x-player

Average loss of x-player's best fixed x vector in hindsight

Jensen's inequality

Average vector of y-player

Concave function: $f(\lambda y + (1-\lambda) y') \ge \lambda f(y) + (1-\lambda) f(y')$



No-Regret Implications

- We know no-regret algorithms exist! (e.g., online gradient descent)

$$x_t = x_{t-1} - \eta \nabla_{\mathrm{x}} \ell(x_{t-1}, y_{t-1}), \qquad y_t = y_{t-1} + \eta \nabla_y \ell(x_{t-1}, y_{t-1})$$

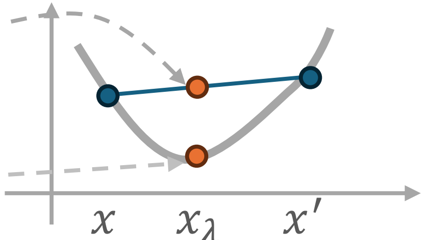
- What this implies is that in the limit as $T \to \infty$ for a regret $\epsilon \to 0$

$$\frac{1}{T}\sum_{t=1}^{T} \ell(x_t, y_t) \le \min_x \frac{1}{T}\sum_{t=1}^{T} \ell(x, y_t) + \epsilon \le \min_x \ell(x, \bar{y}) + \epsilon$$

Jensen's inequality

$$\frac{1}{T}\sum_{t=1}^{T} \ell(x_t, y_t) \ge \max_y \frac{1}{T}\sum_{t=1}^{T} \ell(x_t, y) - \epsilon \ge \max_y \ell(\bar{x}, y) - \epsilon$$

Convex function: $f(\lambda y + (1-\lambda) y') \le \lambda f(y) + (1-\lambda) f(y')$



No-Regret Implications

- We know no-regret algorithms exist! (e.g., online gradient descent)

$$x_t = x_{t-1} - \eta \nabla_{\mathrm{x}} \ell(x_{t-1}, y_{t-1}), \qquad y_t = y_{t-1} + \eta \nabla_y \ell(x_{t-1}, y_{t-1})$$

- What this implies is that in the limit as $T \to \infty$ for a regret $\epsilon \to 0$

$$\boxed{\frac{1}{T}\sum_{t=1}^{T} \ell(x_t, y_t) \le \min_x \ell(x, \bar{y}) + \epsilon}$$

Expected average loss of x-player is a "best-response" to average strategy $\bar{y}$ of y-player

$$\boxed{\frac{1}{T}\sum_{t=1}^{T} \ell(x_t, y_t) \ge \max_y \ell(\bar{x}, y) - \epsilon}$$

Expected average utility of y-player is a "best-response" to average strategy $\bar{x}$ of x-player

No-Regret Implications

- What this implies is that in the limit as $T \to \infty$ for a regret $\epsilon \to 0$

$$\frac{1}{T}\sum_{t=1}^{T} \ell(x_t, y_t) \leq \min_x \ell(x, \bar{y}) + \epsilon \leq \ell(\bar{x}, \bar{y}) + \epsilon$$

$$\Vert$$

$$\frac{1}{T}\sum_{t=1}^{T} \ell(x_t, y_t) \geq \max_y \ell(\bar{x}, y) - \epsilon \geq \ell(\bar{x}, \bar{y}) - \epsilon$$

Average loss of x-player =
Average utility of y-player

$\leq$ $\geq$

Loss (utility) of
average strategies

No-Regret Implications

- What this implies is that in the limit as $T \to \infty$ for a regret $\epsilon \to 0$

$$\ell(\bar{x}, \bar{y}) \geq \max_y \ell(\bar{x}, y) - 2\epsilon$$

$$\ell(\bar{x}, \bar{y}) \leq \min_x \ell(x, \bar{y}) + 2\epsilon$$

$(\bar{x}, \bar{y})$ is a 2ϵ-approximate equilibrium

$(\bar{x}, \bar{y}) \to$ equilibrium as $T \to \infty$

$$\frac{1}{T}\sum_{t=1}^{T} \ell(x_t, y_t) \leq \min_x \ell(x, \bar{y}) + \epsilon \leq \ell(\bar{x}, \bar{y}) + \epsilon$$

$$\|$$

$$\frac{1}{T}\sum_{t=1}^{T} \ell(x_t, y_t) \geq \max_y \ell(\bar{x}, y) - \epsilon \geq \ell(\bar{x}, \bar{y}) - \epsilon$$

Average loss of x-player =
Average utility of y-player

$\leq$
$\geq$

Loss (utility) of
average strategies

Main Takeaway: Equilibrium via No-Regret

Theorem. If two players play repeatedly a *convex-concave zero-sum game* and each player uses any no-regret algorithm to pick their vector (x_t, y_t), then the average vector of each player

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

are a 2ϵ-approximate Nash equilibrium (where ϵ is the regret at of each algorithm after T periods). Hence,

$$(\bar{x}, \bar{y}) \to \text{equilibrium as } T \to \infty$$

Minimax Theorem via No-Regret

- What this implies is that in the limit as $T \to \infty$ for a regret $\epsilon \to 0$

$$\frac{1}{T}\sum_{t=1}^{T} \ell(x_t, y_t) \le \min_x \ell(x, \bar{y}) + \epsilon \le \max_y \min_x \ell(x, y) + \epsilon$$

$$\|$$

$$\frac{1}{T}\sum_{t=1}^{T} \ell(x_t, y_t) \ge \max_y \ell(\bar{x}, y) - \epsilon \ge \min_x \max_y \ell(x, y) - \epsilon$$

Average loss of x-player =
Average utility of y-player

Minimax Theorem via No-Regret

- What this implies is that in the limit as $T \to \infty$ for a regret $\epsilon \to 0$

$$\frac{1}{T}\sum_{t=1}^{T} \ell(x_t, y_t) \leq \min_x \ell(x, \bar{y}) + \epsilon \leq \max_y \min_x \ell(x, y) + \epsilon$$

$$||$$

$$\frac{1}{T}\sum_{t=1}^{T} \ell(x_t, y_t) \geq \max_y \ell(\bar{x}, y) - \epsilon \geq \min_x \max_y \ell(x, y) - \epsilon$$

$$\boxed{\max_y \min_x \ell(x, y) \geq \min_x \max_y \ell(x, y) + 2\epsilon}$$

Average loss of x-player =
Average utility of y-player

Minimax Theorem via No-Regret

Theorem. Existence of no-regret algorithms implies (as $\epsilon \to 0$) that

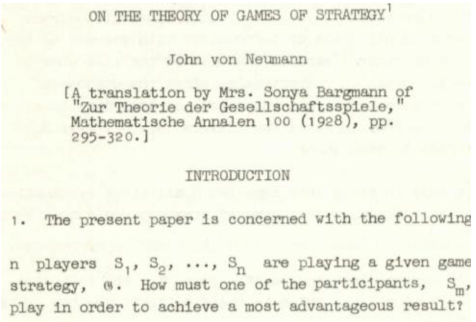
$$\max_{y \in \mathcal{Y}} \min_{x \in \mathcal{X}} \ell(x, y) \geq \min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} \ell(x, y)$$

The other direction is trivial *(why?)*

$$\max_{y \in \mathcal{Y}} \min_{x \in \mathcal{X}} \ell(x, y) \leq \min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} \ell(x, y)$$

Thus

$$\max_{y \in \mathcal{Y}} \min_{x \in \mathcal{X}} \ell(x, y) = \min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} \ell(x, y)$$


ON THE THEORY OF GAMES OF STRATEGY[1]

John von Neumann

[A translation by Mrs. Sonya Bargmann of "Zur Theorie der Gesellschaftsspiele," Mathematische Annalen 100 (1928), pp. 295-320.]

INTRODUCTION

1. The present paper is concerned with the following

n players $S_1, S_2, \ldots, S_n$ are playing a given game strategy, $\mathfrak{S}$. How must one of the participants, S_m, play in order to achieve a most advantageous result?

(an alternative to von Neuman's original proof)

Recap: Equilibrium via No-Regret

Corollary. If two players play repeatedly a zero-sum game, with n rows and m columns, and each player uses EXP with step size $\eta = \sqrt{\log \max(n, m) / 2T}$, to pick their action distributions (x_t, y_t), then the average action distributions of each player

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

are a $2\sqrt{\frac{2\log \max(n,m)}{T}}$ - approximate Nash equilibrium.

Can we do better in terms of rate?

Fast Convergence

- $1/\sqrt{T}$ is tight no-regret rate, if loss sequence chosen by adversary
- When we deploy learning in games, the loss sequence is the outcome of learning of another player
- This is far from adversarial and has many nice properties

- Can we prove faster rates of convergence for learning in games, by leveraging properties of the loss sequence implied by this?

Intuition

- Suppose we use regularized no-regret algorithms (e.g. FTRL)
- Then we know they satisfy stability
$$\|x_t - x_{t-1}\|_1 = O(\eta), \qquad \|y_t - y_{t-1}\|_1 = O(\eta)$$
- The loss of the x-player between two periods is
$$\ell_t = Ay_t, \qquad \ell_{t-1} = Ay_{t-1} \Rightarrow \|\ell_t - \ell_{t-1}\| \leq O(\eta)$$
- Last period loss is very similar to next period loss!

- Can we leverage this fact to device a better no-regret algorithm?

Reminder: FTRL

$$p_t = \operatorname*{argmin}_p \boxed{\sum_{\tau<t} \langle p, \ell_\tau \rangle} + \frac{1}{\eta} \boxed{\mathcal{R}(p)}$$

1-strongly convex function of p that stabilizes the minimizer

Historical performance of always choosing p

$$\mathcal{R}(p) = \sum_{i=1}^{n} p_i \log(p_i) \quad \left(\begin{matrix}\text{Negative}\\ \text{Entropy}\end{matrix}\right)$$

$$p_t \propto p_{t-1} \exp(\eta\, \ell_{t-1})$$

Exponential weight updates algorithm!
(aka Hedge, Multiplicative Weight Updates, EXP, ...)

FTRL with Predictors

Remember Be-the-Leader Lemma: if we know next period loss and play the leader including next period loss, then we have no-regret!

- What if we have a predictor M_t about the next period loss?

- Pretend as if it was the next period loss and play Be-The-Leader

FTRL with Predictors

Predictor of next period loss

$$\binom{\text{FTRL}}{\text{w. Predictors}} \qquad p_t = \underset{p}{\operatorname{argmin}} \boxed{\sum_{\tau<t} \langle p, \ell_\tau \rangle} + \boxed{\langle p, M_t \rangle} + \frac{1}{\eta} \boxed{\mathcal{R}(p)}$$

1-strongly convex function of p that stabilizes the minimizer

Historical performance of always choosing p

$$\boxed{\begin{gathered} \mathcal{R}(p) = \sum_{i=1}^{n} p_i \log(p_i) \quad \binom{\text{Negative}}{\text{Entropy}} \\ p_t \propto p_{t-1} \exp\big(\eta(\ell_{t-1} + M_t - M_{t-1})\big) \end{gathered}}$$

Exponential weight updates with predictors!

Regret of FTRL with Predictors

$$\begin{pmatrix}\text{FTRL} \\ \text{w. Predictors}\end{pmatrix} \qquad p_t = \underset{p}{\operatorname{argmin}} \sum_{\tau<t} \langle p, \ell_\tau \rangle + \langle p, M_t \rangle + \frac{1}{\eta} \mathcal{R}(p)$$

$$(\text{BTRL}) \qquad \tilde{p}_t = \underset{p}{\operatorname{argmin}} \sum_{\tau<t} \langle p, \ell_\tau \rangle + \langle p, \ell_t \rangle + \frac{1}{\eta} \mathcal{R}(p)$$

Theorem. For any loss sequence, with $\ell_t^i \in [0, 1]$:

$$\text{Regret}(\ell_{1:T}) \leq 2 \frac{1}{T} \sum_{t=1}^{T} |\tilde{p}_t - p_t| + \frac{1}{\eta T} \left(\max_p \mathcal{R}(p) - \min_p \mathcal{R}(p) \right)$$

Proof is identical to the bound on FTRL without predictors

How close is FTRL with Predictors to BTRL?

$$\begin{pmatrix}\text{FTRL}\\ \text{w. Predictors}\end{pmatrix} \qquad p_t = \operatorname*{argmin}_p \sum_{\tau<t} \langle p, \ell_\tau \rangle + \langle p, M_t \rangle + \frac{1}{\eta}\mathcal{R}(p)$$

$$(\text{BTRL}) \qquad \tilde{p}_t = \operatorname*{argmin}_p \sum_{\tau<t} \langle p, \ell_\tau \rangle + \langle p, \ell_t \rangle + \frac{1}{\eta}\mathcal{R}(p)$$

Theorem. For the FTRL with predictors: $\|\tilde{p}_t - p_t\|_1 \leq \eta \|\ell_t - M_t\|_\infty$

Proof. Invoke *stability of strongly convex functions theorem* with

$$f(p) = \sum_{\tau<t} \langle p, \ell_\tau \rangle + \langle p, M_t \rangle + \frac{1}{\eta}\mathcal{R}(p), \qquad g(p) = \sum_{\tau<t} \langle p, \ell_\tau \rangle + \langle p, \ell_t \rangle + \frac{1}{\eta}\mathcal{R}(p)$$

$$h(p) = g(p) - f(p) = \langle p, \ell_t - M_t \rangle \Rightarrow \boxed{\|\ell_t - M_t\|_\infty} - \text{Lipschitz w.r.t. } \|\cdot\|_1$$

$\|v\|_\infty = \max_{i=1}^n |v_i|$

How stable is FTRL with Predictors?

$$p_t = \operatorname*{argmin}_p \sum_{\tau<t} \langle p, \ell_\tau \rangle + \langle p, M_t \rangle + \frac{1}{\eta}\mathcal{R}(p)$$
$$p_{t+1} = \operatorname*{argmin}_p \sum_{\tau<t+1} \langle p, \ell_\tau \rangle + \langle p, M_{t+1} \rangle + \frac{1}{\eta}\mathcal{R}(p)$$

Theorem. If losses and predictors lie in $[0,1]^n$: $\|p_{t+1} - p_t\|_1 \leq 3\eta$

Proof. Invoke *stability of strongly convex functions theorem* with

$$f(p) = \sum_{\tau<t} \langle p, \ell_\tau \rangle + \langle p, M_t \rangle + \frac{1}{\eta}\mathcal{R}(p), \qquad g(p) = \sum_{\tau<t+1} \langle p, \ell_\tau \rangle + \langle p, M_{t+1} \rangle + \frac{1}{\eta}\mathcal{R}(p)$$

$$h(p) = g(p) - f(p) = \langle p, M_{t+1} - M_t + \ell_{t+1} \rangle \Rightarrow \boxed{3} - \text{Lipschitz w.r.t. } \|\cdot\|_1$$

Assuming predictors
and losses lie in $[0,1]^n$

Punchline

$$\begin{pmatrix}\text{FTRL} \\ \text{w. Predictors}\end{pmatrix} \qquad p_t = \operatorname*{argmin}_p \underbrace{\boxed{\sum_{\tau<t} \langle p, \ell_\tau \rangle}}_{\text{Historical performance of always choosing } p} + \overbrace{\boxed{\langle p, M_t \rangle}}^{\text{Predictor of next period loss}} + \frac{1}{\eta}\boxed{\mathcal{R}(p)}$$

1-strongly convex function of p that stabilizes the minimizer

Corollary. FTRL with predictors is 3η-stable and has regret

$$\leq \underbrace{\boxed{\frac{\eta}{T}\sum_{t=1}^{T} \|\ell_t - M_t\|_\infty}}_{\text{Average stability with respect to BTRL induced by regularizer}} + \underbrace{\boxed{\frac{1}{\eta T}\left(\max_p \mathcal{R}(p) - \min_p \mathcal{R}(p)\right)}}_{\text{Average loss distortion caused by regularizer}}$$

What is a good predictor in the context of games?

Optimistic FTRL: Last Period Predictor

Optimism: predict that the next period loss will be the same as last period loss

$$\begin{pmatrix}\text{FTRL} \\ \text{w. Predictors}\end{pmatrix} \qquad p_t = \operatorname*{argmin}_p \boxed{\sum_{\tau<t} \langle p, \ell_\tau \rangle} + \boxed{\langle p, \ell_{t-1} \rangle} + \frac{1}{\eta} \boxed{\mathcal{R}(p)}$$

1-strongly convex function of p that stabilizes the minimizer

Historical performance of always choosing p

$$\mathcal{R}(p) = \sum_{i=1}^{n} p_i \log(p_i) \quad \begin{pmatrix}\text{Negative} \\ \text{Entropy}\end{pmatrix}$$

$$p_t \propto p_{t-1} \exp\big(\eta \, (2\ell_{t-1} - \ell_{t-2})\big)$$

Optimistic Exponential Weight Updates!

Optimistic EXP

Corollary. Optimistic EXP is 3η-stable and has regret

$$R(T) \le \frac{\eta}{T} \boxed{\sum_{t=1}^{T} \|\ell_t - \ell_{t-1}\|_\infty} + \frac{\log(n)}{\eta\, T}$$

Average stability of the loss vector

Applying Optimistic EXP to Games

Suppose both players use Optimistic EXP with step-size η

$$R_x(T) \le \frac{\eta}{T}\sum_{t=1}^{T} \underbrace{\boxed{\|A(y_t - y_{t-1})\|_\infty}}_{\text{stability of loss vector}} + \frac{\log(n)}{\eta\, T}$$

$$\le \frac{\eta}{T}\sum_{t=1}^{T} \underbrace{\boxed{\|y_t - y_{t-1}\|_1}}_{\text{stability of opponent}} + \frac{\log(n)}{\eta\, T}$$

$$\le \frac{\eta}{T}\sum_{t=1}^{T} \boxed{3\eta} + \frac{\log(n)}{\eta T} = \boxed{3\eta^2} + \frac{\log(n)}{\eta T}$$

Since opponent uses an η-stable algorithm

Much smaller leading term (closeness to BTRL) than without predictors (i.e. η^2 vs. η)

Optimistic EXP Dynamics

Larger step size than if we were playing against an adversary
$T^{-1/3}$ vs. $T^{-1/2}$
(e.g. if $T = 1000$, then 0.1 vs. 0.032)

Corollary. If all players use Optimistic EXP with $\boxed{\eta = \left(\frac{\log(n \vee m)}{T}\right)^{1/3}}$ then each player's regret is at most $\boxed{\epsilon = 4\left(\frac{\log(n \vee m)}{T}\right)^{2/3}}$ and the average vectors $(\bar{x}, \bar{y})$ are an 2ϵ-approximate equilibrium

Order of magnitude smaller regret than playing against an adversary
$T^{-2/3}$ vs. $T^{-1/2}$
(e.g. if $T = 1000$, then 0.01 vs. 0.032)

Optimistic EXP Dynamics

An even better theorem can be proven with a more refined analysis

[1311.1869] Optimization, Learning, and Games with Predictable Sequences (arxiv.org)

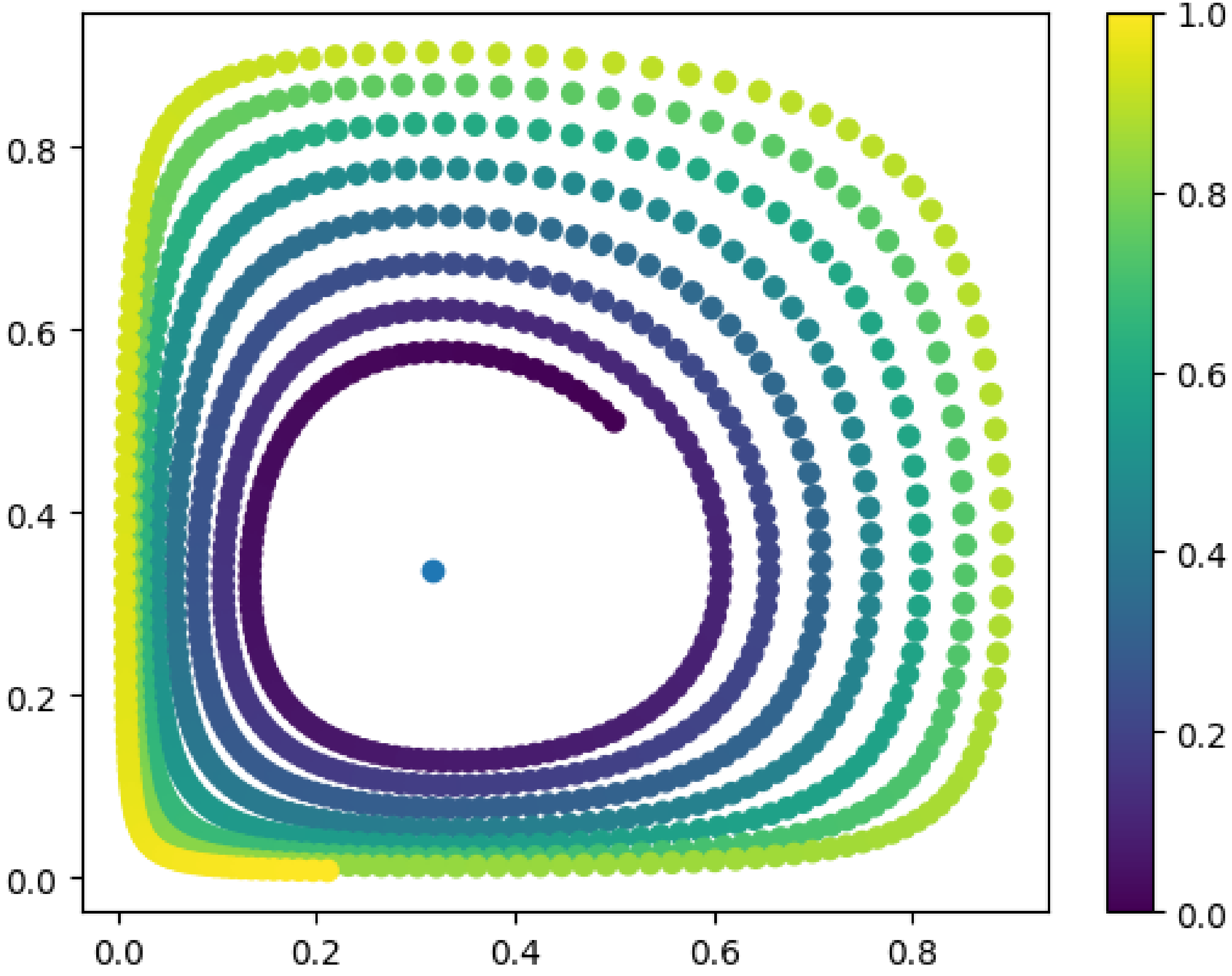
Theorem [Rakhlin-Sridharan'13]. If players use Optimistic EXP with $\eta = O(1)$ then the average vectors $(\bar{x}, \bar{y})$ are an $O\left(\frac{\log(n \vee m)}{T}\right)$-approximate equilibrium.

Intuition. Utilizes the fact that $\epsilon = R_x + R_y$. One can prove bounds on R_x that contain more refined "negative terms" (typically ignored). Rather than ignoring them, these negative terms cancel out with positive terms in R_y, when you sum the two regret terms.

Do the dynamics actually converge?

$(\bar{x}, \bar{y}) \to \text{equilibrium}$ vs. $(x_T, y_T) \to \text{equilibrium}$

"average iterate convergence" "last-iterate convergence"



A simple example

Consider the game defined by loss matrix

$$A = \begin{pmatrix} .5 & 0 \\ 0 & 1 \end{pmatrix}$$

EXP dynamics:

$$x_t \propto x_{t-1} \exp(-\eta A y_{t-1})$$

$$y_t \propto y_{t-1} \exp\left(\eta A^\top x_{t-1}\right)$$

A Simple Game Analysis

- Consider the simplest convex-concave zero-sum game

$$\ell(x, y) = xy, \qquad x \in R, y \in R$$

- The only equilibrium of this game is $(0, 0)$ (why?)
- What if both player use online gradient descent

$$x_t = x_{t-1} - \eta \nabla_x \ell(x_{t-1}, y_{t-1}) = x_{t-1} - \eta y_{t-1}$$
$$y_t = y_{t-1} + \eta \nabla_y \ell(x_{t-1}, y_{t-1}) = y_{t-1} + \eta x_{t-1}$$

- What happens to the distance to equilibrium at each period

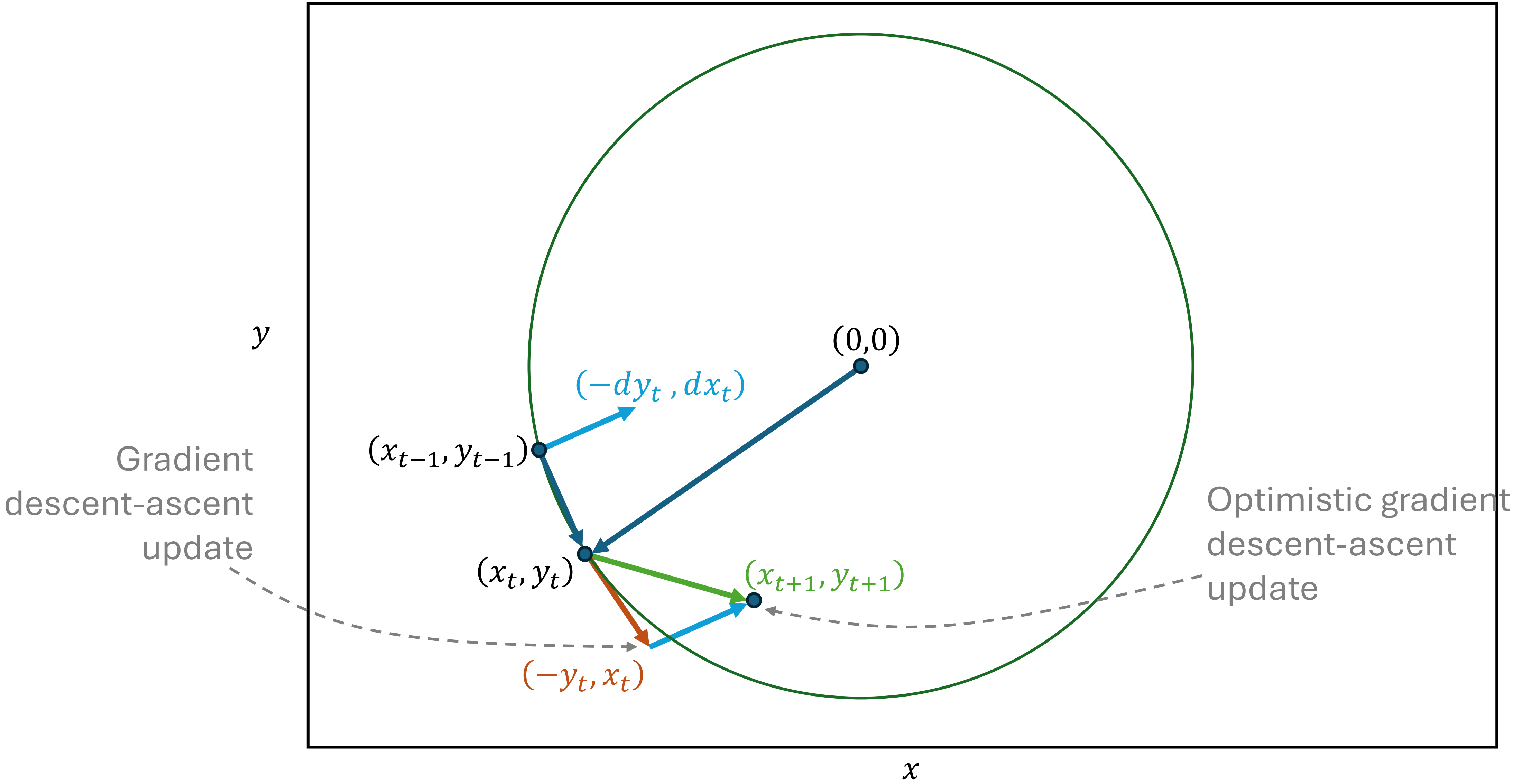
$$\begin{aligned} x_t^2 + y_t^2 &= x_{t-1}^2 \underline{- 2\eta x_{t-1} y_{t-1}} + \eta^2 y_{t-1}^2 + y_{t-1}^2 \underline{+ 2\eta x_{t-1} y_{t-1}} + \eta^2 x_{t-1}^2 \\ &= \left(1 + \eta^2\right)\left(x_{t-1}^2 + y_{t-1}^2\right) \end{aligned}$$

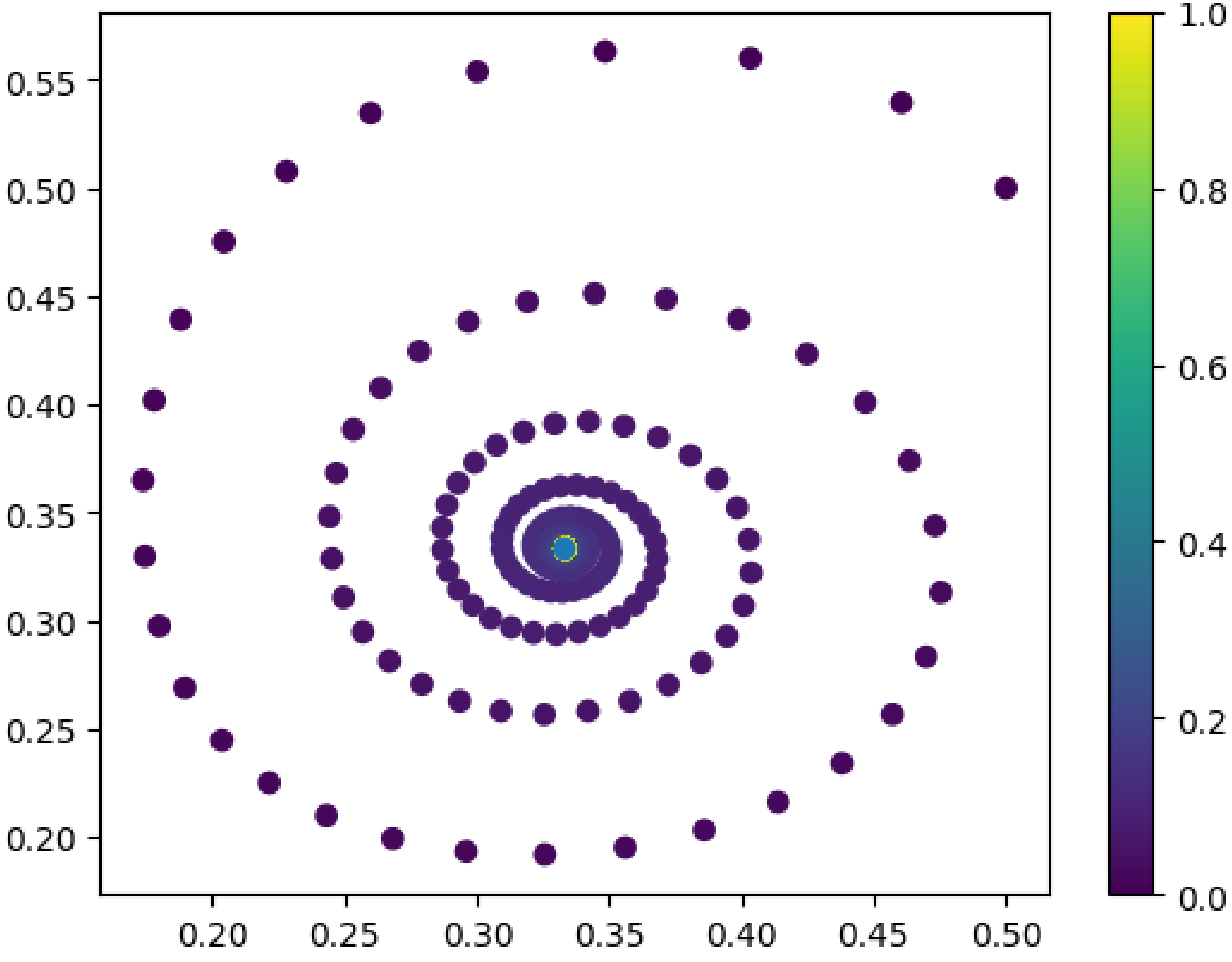
- It grows!! We move away from equilibrium

A Simple Game Analysis

- Consider the simplest convex-concave zero-sum game
$$\ell(x, y) = xy, \qquad x \in R, y \in R$$
- The only equilibrium of this game is $(0, 0)$ (why?)
- What if both player use optimistic online gradient descent
$$x_t = x_{t-1} - \eta(2y_{t-1} - y_{t-2}) = x_{t-1} - \eta y_{t-1} - \eta(y_{t-1} - y_{t-2})$$
$$y_t = y_{t-1} + \eta(2x_{t-1} - x_{t-2}) = y_{t-1} + \eta x_{t-1} + \eta(x_{t-1} - x_{t-2})$$
- What happens to the distance to equilibrium at each period?
$$dx_t \coloneqq x_t - x_{t-1} = -\eta y_{t-1} - \eta dy_{t-1}$$
$$dy_t \coloneqq y_t - y_{t-1} = \quad \eta x_{t-1} + \eta dx_{t-1}$$
A form of "negative momentum"

A form of negative momentum





A simple example

Consider the game defined by loss matrix

$$A = \begin{pmatrix} .5 & 0 \\ 0 & 1 \end{pmatrix}$$

Optimistic EXP dynamics:

$$x_t \propto x_{t-1} \exp\big(-\eta(2Ay_{t-1} - Ay_{t-2})\big)$$

$$y_t \propto y_{t-1} \exp\left(\eta\big(2A^\top x_{t-1} - A^\top x_{t-2}\big)\right)$$

Convergence of Optimistic EXP

- Define *distance to equilibrium* as the KL-divergence:

$$d_t \coloneqq KL\big((x_t, y_t)\|(x_*, y_*)\big) = \left\langle x_*, \log\left(\frac{x_*}{x_t}\right)\right\rangle + \left\langle y_*, \log\left(\frac{y_*}{y_t}\right)\right\rangle$$

- We will investigate whether this distance decreases at each period:

$$\Delta_t \coloneqq d_{t+1} - d_t = -\left\langle x_*, \log\left(\frac{x_{t+1}}{x_t}\right)\right\rangle - \left\langle y_*, \log\left(\frac{y_{t+1}}{y_t}\right)\right\rangle$$

Theorem. For η smaller than some constant, when (x_t, y_t) is "$\Omega\left(\eta^{1/3}\right)$*-far from* (x_*, y_*)" then

$$\Delta_t \leq -\Omega\left(\eta^3\right)$$

Hence, eventually (x_t, y_t) will be "$O\left(\eta^{1/3}\right)$*-close to* (x_*, y_*)"

Appendix

Main arguments in proof of convergence of Optimistic EXP

[1807.04252] Last-Iterate Convergence: Zero-Sum Games and Constrained Min-Max Optimization (arxiv.org)

Convergence of Optimistic EXP

Coordinate-wise multiplication of two vectors

$$x_{t+1} = \frac{x_t \cdot \exp(-2\eta A y_t + \eta A y_{t-1})}{\langle x_t, \exp(-2\eta A y_t + \eta A y_{t-1})\rangle}, \qquad y_{t+1} = \frac{y_t \cdot \exp\left(2\eta A^\top x_t - \eta A^\top x_{t-1}\right)}{\langle y_t, \exp(2\eta A^\top x_t - \eta A^\top x_{t-1})\rangle}$$

- Decrease in distance simplifies to:

$$\begin{aligned}\Delta_t = {} & \boxed{\langle x_*, \eta A(2y_t - y_{t-1})\rangle - \left\langle y_*, \eta A^\top (2x_t - x_{t-1})\right\rangle} \\ & + \log\langle x_t, \exp(-2\eta A y_t + \eta A y_{t-1})\rangle \\ & + \log\left\langle y_t, \exp\left(2\eta A^\top x_t - \eta A^\top x_{t-1}\right)\right\rangle\end{aligned}$$

- **First part ≤ 0.** For small η, $2y_t - y_{t-1}$ and $2x_t - x_{t-1}$ lie in simplices. By equilibrium:

$$x_*^\top A y_* \leq (2x_t - x_{t-1})^\top A y_*, \qquad x_*^\top A y_* \geq x_*^\top A(2y_t - y_{t-1})$$

- **Second part.** Use Taylor approximations and definition of dynamics

Convergence of Optimistic EXP

$$\Delta_t \leq \log\langle x_t, \exp(-2\eta A y_t + \eta A y_{t-1})\rangle + \log\langle y_t, \exp\left(2\eta A^\top x_t - \eta A^\top x_{t-1}\right)\rangle$$

- Both quantities can be viewed as a weighted soft-max operator over a vector
- We will consider a Taylor approximation to the softmax after centering
- For simplicity define $v_t = A(2y_t - y_{t-1})$ and $u_t = A^\top(2x_t - x_{t-1})$, so that

$$\boxed{\Delta_t \leq \log\langle x_t, \exp(-\eta v_t)\rangle + \log\langle y_t, \exp(\eta u_t)\rangle}$$

Side note: if we were to use EXP then we can derive the same bound but with $v_t = A y_t$ and $u_t = A^\top x_t$

- Center vectors around scalars $\bar{v}_t, \bar{u}_t$, so that average deviations from centers are "small"

$$\Delta_t \leq -\eta \bar{v}_t + \log\langle x_t, \exp\left(-\eta(v_t - \bar{v}_t)\right)\rangle + \eta \bar{u}_t + \log\langle y_t, \exp\left(\eta(u_t - \bar{u}_t)\right)\rangle$$

Convergence of Optimistic EXP

$$-\eta\bar{v}_t + \log\langle x_t, \exp\underbrace{\boxed{\left(-\eta(v_t - \bar{v}_t)\right)}}^{r_t}\rangle$$

- Consider a second order Taylor approximation to "exp"

$$\log\left\langle x_t, 1 + r_t + \left(\frac{1}{2} + O(\eta)\right) r_t^2 \right\rangle = \log\left(1 + \langle x_t, r_t\rangle + \left(\frac{1}{2} + O(\eta)\right)\langle x_t, r_t^2\rangle\right)$$

- Choose centers so that the first order term vanishes (i.e., $\bar{v}_t = x_t^\top v_t$ and $\bar{u}_t = y_t^\top u_t$)

$$\langle x_t, r_t\rangle = -\eta\left(x_t^\top v_t - \bar{v}_t\right) = 0$$

- For the second order, we can simply upper bound using $\log(1+x) \leq x$

$$-\eta\, x_t^\top v_t + \left(\frac{1}{2} + O(\eta)\right)\eta^2\langle x_t, (v_t - \langle x_t, v_t\rangle)^2\rangle$$

Convergence of Optimistic EXP

Decrease in distance is upper bounded by

$$\Delta_t \le \eta\, y_t^\top u_t - \eta\, x_t^\top v_t + \left(\frac{1}{2} + O(\eta)\right)\eta^2 \Big(\overbrace{\langle x_t, (v_t - \langle x_t, v_t\rangle)^2 \rangle}^{R_t^x} + \overbrace{\langle y_t, (u_t - \langle y_t, u_t\rangle)^2 \rangle}^{R_t^y} \Big)$$

Quantity $v_t - \langle x_t, v_t\rangle$ can be thought as a mixture of "regrets" of each action of x-player

$$v_t - \langle x_t, v_t\rangle = 2\big(Ay_t - x_t^\top A y_t\big) - \big(Ay_{t-1} - x_t^\top A y_{t-1}\big)$$

Definition. We say that a point is $\eta^{1/3}$ far from equilibrium if at least one entry with weight $x_t^i = \Omega\big(\eta^{1/3}\big)$ has regret $x_t^\top A y_t - (Ay_t)_i = \Omega\big(\eta^{1/3}\big)$

Given that algorithm is η-stable, we also have that $\|y_t - y_{t-1}\| \le O(\eta)$

$$|(v_t - \langle x_t, v_t\rangle)_i| = \left| -2\big(Ay_t - x_t^\top A y_t\big)_i + \big(Ay_{t-1} - x_t^\top A y_{t-1}\big)_i \right| = \Omega\big(\eta^{1/3}\big) - O(\eta) = \Omega\big(\eta^{1/3}\big)$$

Corollary. If we are $\eta^{1/3}$-far from equilibrium then $\max\{R_t^x, R_t^y\} = \Omega(\eta)$

Convergence of Optimistic EXP

Decrease in distance is upper bounded by

$$\Delta_t \leq \eta\, y_t^\top u_t - \eta\, x_t^\top v_t + \left(\frac{1}{2} + O(\eta)\right)\eta^2\left(R_t^x + R_t^y\right)$$

Suppose we can also argue the following main lemma

Main Lemma. $y_t^\top u_t - x_t^\top v_t \leq -(1 - O(\eta))\eta\,\left(R_t^x + R_t^y\right) + O\left(\eta^2\right)$

Combined with the corollary in the previous slide, we get the main theorem

$$\Delta_\mathrm{t} \leq -\left(\frac{1}{2} - O(\eta)\right)\eta^2 \max\left\{R_t^x, R_t^y\right\} + O\left(\eta^3\right) \leq -\Omega\left(\eta^3\right)$$

Main Lemma. $y_t^\top u_t - x_t^\top v_t \le -(1 - O(\eta))\eta \max\{R_t^x, R_t^y\} + O(\eta^2)$

$$y_t^\top u_t - x_t^\top v_t = 2x_t^\top A y_t - x_{t-1}^\top A y_t - 2x_t^\top A y_t + x_t^\top A y_{t-1} = x_t^\top A y_{t-1} - x_{t-1}^\top A y_t$$

- Note that by adding and subtracting the previous period utility/loss:

$$x_t^\top A y_{t-1} - x_{t-1}^\top A y_t = x_t^\top A y_{t-1} - \frac{1}{2} x_{t-1}^\top A y_{t-1} + \frac{1}{2} x_{t-1}^\top A y_{t-1} - x_{t-1}^\top A y_t = \frac{1}{2} y_{t-1}^\top u_t - \frac{1}{2} x_{t-1}^\top v_t$$

- We can derive that:

$$\boxed{y_t^\top u_t - x_t^\top v_t = \frac{1}{2} y_{t-1}^\top u_t - \frac{1}{2} x_{t-1}^\top v_t}$$

- Suppose that we can argue that

(Main Sub-Lemma)
$$x_t^\top v_t - x_{t-1}^\top v_t \le -(1 - O(\eta))\eta R_t^x + O(\eta^2)$$
$$y_{t-1}^\top u_t - y_t^\top u_t \le -(1 - O(\eta))\eta R_t^y + O(\eta^2)$$

- Then $y_t^\top u_t - x_t^\top v_t = \frac{1}{2}(y_t^\top u_t - x_t^\top v_t) - \left(\frac{1}{2} - O(\eta)\right)\eta \max\{R_t^x, R_t^y\}$

- Rearranging yields the lemma

This wouldn't be the case under EXP, where $v_t = A y_t$ and $u_t = A^\top x_t$ in which case $y_t^\top u_t - x_t^\top v_t = 0$.

For optimistic EXP this difference is the bias that shrinks us towards the equilibrium.

Main Sub-Lemma. $x_t^\top v_t - x_{t-1}^\top v_t \leq -\big(1 - O(\eta)\big)\eta R_t^x + O\big(\eta^2\big)$

Suffices to argue lemma for first-order approx. to the Optimistic EXP updates

$$\tilde{x}_t = \frac{x_{t-1} \cdot (1 - \eta v_{t-1})}{\langle x_{t-1}, 1 - \eta v_{t-1} \rangle}$$

Since, it can be argued that first-order approx. is close to original variant, i.e.

$$\|x_t - \tilde{x}_t\| = O\big(\eta^2\big)$$

Thus, we want

$$\tilde{x}_t^\top v_t - x_{t-1}^\top v_t = -\big(1 - O(\eta)\big)\,\eta\,\langle x_t, (v_t - \langle x_t, v_t\rangle)^2\rangle$$

Further since $\|x_t - x_{t-1}\| = O(\eta)$, it suffices that:

$$\tilde{x}_t^\top v_t - x_{t-1}^\top v_t = -\big(1 - O(\eta)\big)\,\eta\,\langle x_{t-1}, (v_t - \langle x_{t-1}, v_t\rangle)^2\rangle$$

Main Sub-Lemma: $x_t^\top v_t - x_{t-1}^\top v_t \le -(1 - O(\eta))\eta R_t^x + O(\eta^2)$

Let $v' = A(2y_{t-1} - y_{t-2}), v = A(2y_t - y_{t-1}), x = x_{t-1}$, and $\tilde{x} = \tilde{x}_t$. We want to show that

$$\langle \tilde{x}, v\rangle - \langle x, v\rangle = -(1 - O(\eta))\eta \left\langle x, (v - \langle x, v\rangle)^2\right\rangle, \qquad \tilde{x} = \frac{x\,(1 - \eta v')}{1 - \eta\langle x, v'\rangle}$$

- Plugging in the update rule for $\tilde{x}$ and simplifying

$$\langle \tilde{x}, v\rangle - \langle x, v\rangle = \frac{\langle x, v\rangle}{1 - \eta\langle x, v'\rangle} - \eta\,\frac{\langle x \cdot v', v\rangle}{1 - \eta\langle x, v'\rangle} - \langle x, v\rangle = \frac{\eta\langle x, v\rangle\langle x, v'\rangle}{1 - \eta\langle x, v'\rangle} - \frac{\eta\langle x, v \cdot v'\rangle}{1 - \eta\langle x, v'\rangle}$$

- By stability $\|v - v'\| = O(\eta)$ and we can derive

$$\langle \tilde{x}, v\rangle - \langle x, v\rangle = \frac{\eta\left(\langle x, v\rangle^2 - \left\langle x, v^2\right\rangle\right)}{1 - \eta\langle x, v'\rangle} + O\left(\eta^2\right)$$

- Note $\left\langle x, (v - \langle x, v\rangle)^2\right\rangle$ is variance of the vector v under distribution x. By variance formula

$$\left\langle x, (v - \langle x, v\rangle)^2\right\rangle = \left\langle x, v^2\right\rangle - \langle x, v\rangle^2$$

- Since $1 - \eta\langle x, v'\rangle \le 1 + O(\eta)$:

$$\langle \tilde{x}, v\rangle - \langle x, v\rangle = -\frac{\eta\left\langle x, (v - \langle x, v\rangle)^2\right\rangle}{1 - \eta\langle x, v'\rangle} + O\left(\eta^2\right) \le -(1 - O(\eta))\,\eta\,\left\langle x, (v - \langle x, v\rangle)^2\right\rangle + O\left(\eta^2\right)$$

Punchline: Last-Iterate Convergence to Equilibrium

For η small enough, when (x_t, y_t) is not $O\left(\eta^{1/3}\right)$-close to (x_*, y_*)

$$\Delta_{\mathrm{t}} \coloneqq d_{t+1} - d_t \leq -\Omega\left(\eta^3\right)$$

Thus eventually (x_t, y_t) will be $\eta^{1/3}$-close to (x_*, y_*).

Some technicalities are also required to show that the definition of closeness used in the proof, also imply closeness with more standard definitions like ℓ_1 distance.