Optimistic Regret Minimization for Extensive-Form Games via Dilated Distance-Generating Functions

Gabriele Farina[1] Christian Kroer[2]
Tuomas Sandholm[1,3,4,5]

[1] Computer Science Department, Carnegie Mellon University
[2] IEOR Department, Columbia University
[3] Strategic Machine, Inc.
[4] Strategy Robot, Inc.
[5] Optimized Markets, Inc.

Outline

- Part 1: Foundations
  - Bilinear saddle-point problems
  - Regret minimization and relationship with saddle points
- Part 2: Recent Advances --- optimistic regret minimization
  - Accelerated convergence to saddle points
  - Example of optimistic/predictive regret minimizers
- Part 3: Applications to game theory
  - Extensive-form games (EFGs)

Contributions

- How to instantiate optimistic regret minimizers in EFGs
- Comparison to non-optimistic methods in extensive-form games
- Experimental observations

Part 1: Foundations

- Bilinear saddle-point problems
- Regret minimization

Bilinear Saddle-Point Problems

- Optimization problems of the form

$$\min_{x \in X} \max_{y \in Y} x^T A y$$

  where X and Y are convex and compact sets, and A is a real matrix.
- Ubiquitous in game theory:
  - Nash equilibrium in zero-sum games
  - Trembling-hand perfect equilibrium
  - Correlated equilibrium, etc.

Bilinear Saddle-Point Problems

- Quality metric: **saddle-point gap**
- Gap of approximate solution (x, y):

$$\xi(x, y) \coloneqq \max_{y' \in Y} x^T A y' - \min_{x' \in X} (x')^T A y$$

- In the context of approximate Nash equilibrium, the gap represents the “exploitability” of the strategy profile

Regret Minimization

- Regret minimizer: device for repeated decision making that supports two operations
  - It outputs the next decision, $x^{t+1} \in X$
  - It receives/observes a linear loss function ℓ^t used to evaluate the last decision, x^t
- The learning is **online**, in the sense that the next decision x^{t+1} is based only on the previous decision $x^1, \dots, x^t$ and corresponding observed losses $\ell^1, \dots, \ell^t$
  - **No assumption available on future losses!**
  - **Must handle adversarial environments**

Regret Minimization

- Quality metric for the device: **cumulative regret**

  "How well do we do against best fixed decision in hindsight?"

$$R^T := \sum_{t=1}^{T} \ell^t(x^t) - \min_{\hat{x} \in X} \left\{ \sum_{t=1}^{T} \ell^t(\hat{x}) \right\}$$

- Goal: make sure that the regret grows at a sublinear rate
  - Many general-purpose regret minimizers known in the literature achieve $O(\sqrt{T})$ cumulative regret
  - This matches the learning-theoretic bound of $\Omega(\sqrt{T})$

Regret Minimization

- Quality metric for the device: **cumulative regret**

"How well do we do against best fixed decision in hindsight?"

$$R^T := \sum_{t=1}^{T} \ell^t(x^t) - \min_{\hat{x} \in X} \left\{ \sum_{t=1}^{T} \ell^t(\hat{x}) \right\}$$

Connection with Saddle Points

- Regret minimization can be used to converge to saddle-point
  - Great success in game theory (e.g., Libratus)

Connection with Saddle Points

- Regret minimization can be used to converge to saddle-point
  - Great success in game theory (e.g., Libratus)
- Take the bilinear saddle-point problem $\min_{x \in X} \max_{y \in Y} x^T A y$
  - Instantiate a regret minimizer for set X and one for set Y
  - At each time t, the regret minimizer for X observes loss Ay^t
  - … and the regret minimizer for Y observes loss $-A^T x^t$

"Self-play"

Connection with Saddle Points

- Regret minimization can be used to converge to saddle-point
  - Great success in game theory (e.g., Libratus)
- Take the bilinear saddle-point problem $\min_{x \in X} \max_{y \in Y} x^T A y$
  - Instantiate a regret minimizer for set X and one for set Y
  - At each time t, the regret minimizer for X observes loss Ay^t
  - … and the regret minimizer for Y observes loss $-A^T x^t$

"Self-play"

- Well-known folk lemma: at each time T, the profile of average decisions $(\bar{x}, \bar{y})$ produced by the regret minimizers has gap

$$\xi(\bar{x}, \bar{y}) \leq \frac{R_X^T + R_Y^T}{T} = O\left(\frac{1}{\sqrt{T}}\right)$$

Recap of Part 1

- Saddle-point problems are min-max problems over convex sets
  - Many game-theoretical equilibria can be expressed as saddle-point problems, including Nash equilibrium
- Regret minimization is a powerful paradigm in online convex optimization
  - Useful to converge to saddle-points in "self-play"
  - Assumes no information is available on the future loss
  - Optimal convergence rate (in terms of saddle-point gap): $\Theta\left(\frac{1}{\sqrt{T}}\right)$

Part 2: Recent Advances (Optimistic/predictive regret minimization)

- Examples of optimistic regret minimizers
- Accelerated convergence to saddle points

Optimistic/Predictive Regret Minimization

- Recent breakthrough in online learning
- Similar to regular regret minimization
- Before outputting each decision x^t, the predictive regret minimizer also receives a **prediction** m^t of the (next) loss function ℓ^t
  - Idea: the regret minimizer should take advantage of this prediction to produce better decisions
  - Requirement: a predictive regret minimizer must guarantee that the regret **will not grow** should the predictions be always correct

Required Regret Bound

- Enhanced requirement on regret growth

$$R^T \leq \alpha + \beta \sum_{t=1}^{T} \|\ell^t - m^t\|_*^2 - \gamma \sum_{t=1}^{T} \|x^t - x^{t-1}\|_*^2$$

Required Regret Bound

- Enhanced requirement on regret growth

$$R^T \leq \alpha + \beta \sum_{t=1}^{T} \|\ell^t - m^t\|_*^2 - \gamma \sum_{t=1}^{T} \|x^t - x^{t-1}\|_*^2$$

Penalty for wrong predictions

Required Regret Bound

- Enhanced requirement on regret growth

$$R^T \leq \alpha + \beta \boxed{\sum_{t=1}^{T} \|\ell^t - m^t\|_*^2} - \gamma \sum_{t=1}^{T} \|x^t - x^{t-1}\|_*^2$$

Penalty for wrong predictions

- Predictive regret minimizers exist
  - Optimistic follow-the-regularized leader (Optimistic FTRL) [Syrgkanis et al., 2015]
  - Optimistic online mirror descent (Optimistic OMD) [Rakhlin and Sridharan, 2013]

FTRL

- Picks the next decision x^{t+1} according to

$$x^{t+1} = \operatorname{argmin}_{x \in X} \left\langle \sum_{\tau=1}^{t} \ell^{\tau}, x \right\rangle + \frac{1}{\eta} d(x),$$

where $d(x)$ is a **1-strongly convex regularizer** over X.

Optimistic FTRL

- Picks the next decision x^{t+1} according to

$$x^{t+1} = \operatorname{argmin}_{x \in X} \left\langle m^{t+1} + \sum_{\tau=1}^{t} \ell^{\tau}, x \right\rangle + \frac{1}{\eta} d(x),$$

where $d(x)$ is a **1-strongly convex regularizer** over X.

Optimistic OMD

- Slightly more complicated rule for picking the next decision
- Implementation again parametric on a 1-strongly convex regularizer just like optimistic FTRL

Accelerated convergence to saddle points

- When the prediction m^t is set up to be equal to ℓ^{t-1}, one can improve the folk lemma:

  The average decisions output by predictive regret minimizers that face each other satisfy

$$\xi(\bar{x}, \bar{y}) = O\left(\frac{1}{T}\right)$$

  – This again matches the learning-theoretic bound for (accelerated) first-order methods

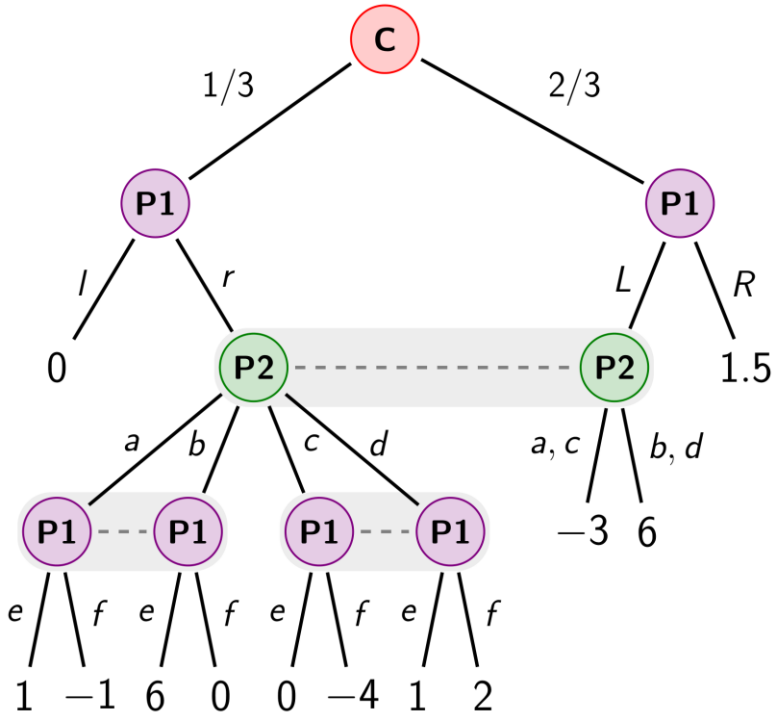
Recap of Part 2

- Predictive regret minimization is a recent breakthrough in online learning
- Idea: predictive regret minimizers receive a prediction of the next loss
- “Good” predictive regret minimizers exist in the literature
- Predictive regret minimizers enable to break the learning theoretic bound of $\Theta\left(\frac{1}{\sqrt{T}}\right)$ convergence to saddle points, and enable accelerated $\Theta\left(\frac{1}{T}\right)$ convergence instead.

Part 3: Applications to Game Theory

- Extensive-form games
- How to construct regularizers in games

Extensive-Form Games



- Can capture sequential and simultaneous moves
- Private information
- Each information set contains a set of “undistinguishable” tree nodes
  - Information sets correspond to **decision points** in the game
- We assume perfect recall: no player forgets what the player knew earlier

Decision Space for an Extensive-Form Game

- The set of strategies in an extensive-form games is best expressed in **sequence form** [von Stengel, 1996]
  - For each action a at decision point/information set j, associate a real number that represents the probability of the player taking all actions on the path from the root of the tree to that (information set, action) pair
- (Non-predictive) regret minimizers that can output decisions on the space of sequence-form strategies exist
  - Notably, CFR and its later variants CFR+ [Tammelin et al., 2015] and Linear CFR [Brown and Sandholm, 2019]
  - Great practical success, but suboptimal $O\left(\frac{1}{\sqrt{T}}\right)$ convergence rate to equilibrium

Natural Question

How can we set up optimistic regret minimizers for the space of sequence-form strategies?

Regularizers for Sequence-Form Strategies

- Both optimistic FTRL and optimistic OMD are parametric on a choice of regularizers for the domain of decisions
  - In the case of extensive-form games: space of sequence-form strategies
- In the paper we focus on **dilated** regularizers:
  - Pick a **local regularizer** at each decision point in the game
  - "Connect" the local regularizer via **dilation** (a convexity-preserving operation)

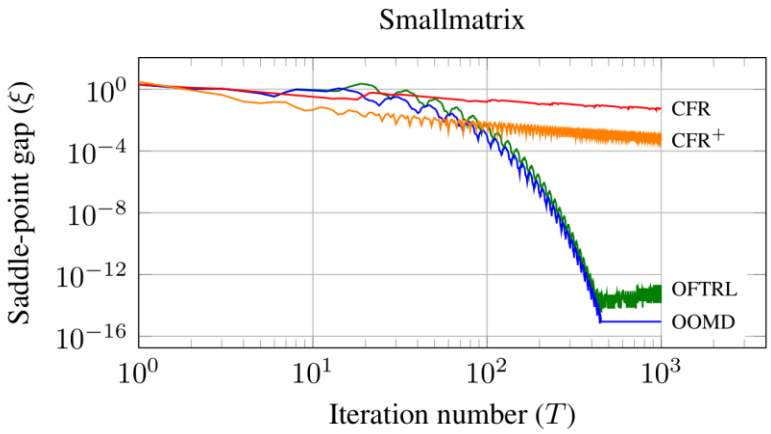
Regularizers for Sequence-Form Strategies

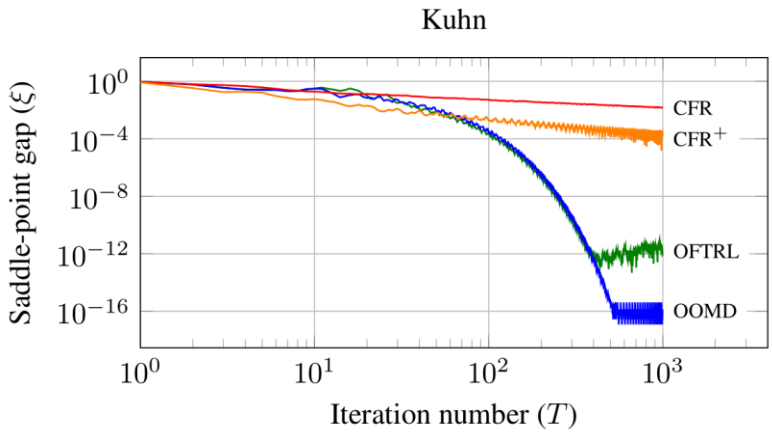
- We give a framework for how to set up dilated regularizers in extensive-form games
- We give guarantees on the strong convexity modulus of the regularizers (wrt Euclidean norm)
- We give specific examples of such regularizers
- These regularizers can be used in conjunction with optimistic FTRL and optimistic OMD to converge to equilibrium as $\Theta\left(\frac{1}{T}\right)$

Dilated Regularizers Imply Local Regret Minimization

- We show that optimistic FTRL and optimistic OMD instantiated with our regularizers decompose regret over the extensive-form strategy space as a sum of contributions local to each information set
- Optimistic OMD in particular can be seen as using local regret minimizers, one for each information set, to minimize regret over the whole sequential strategy space
- This matches the CFR paradigm, the leading state of the art in extensive-form game solving

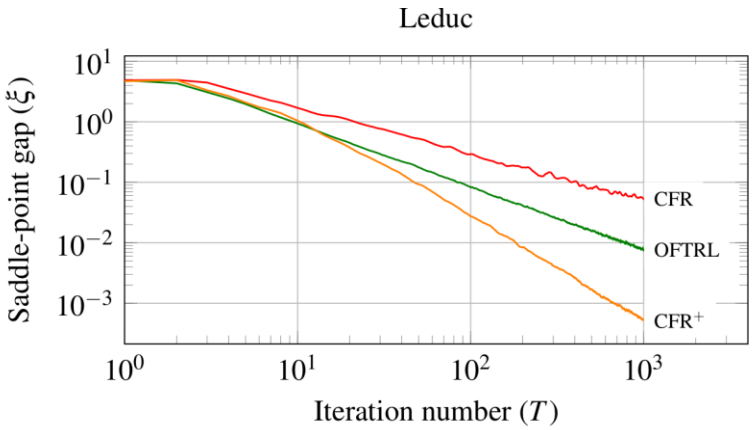
Experimental Observations

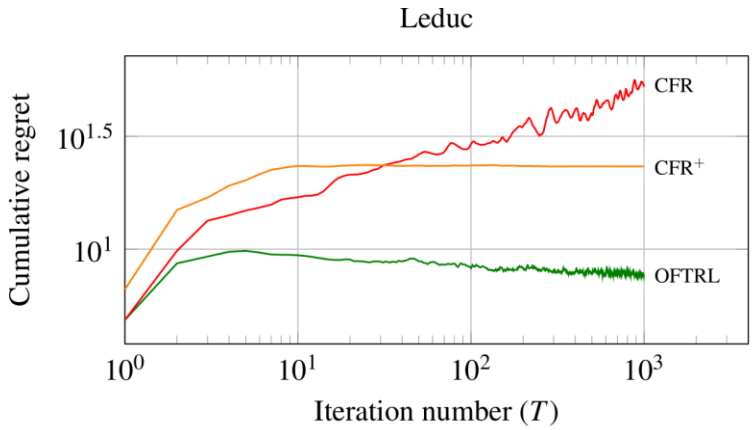




- Several orders of magnitude faster than CFR/CFR+ in shallow games

Experimental Observations





- On the other hand, deeper games seem to pose more challenges

Conclusions

- We studied how optimistic regret minimization can be applied in the context of extensive-form games
  - Fundamental ingredient: tractable regularizers for the domain at hand (extensive-form strategy space)
- First explicit bound on strong convexity properties of dilated distance-generating functions wrt Euclidean norm
- We prove that regret updates are local at each decision point
- In shallow games, these methods can outperform state-of-the-art CFR/CFR+ by up to 12 orders of magnitude
  - Acceleration in deeper games remains elusive